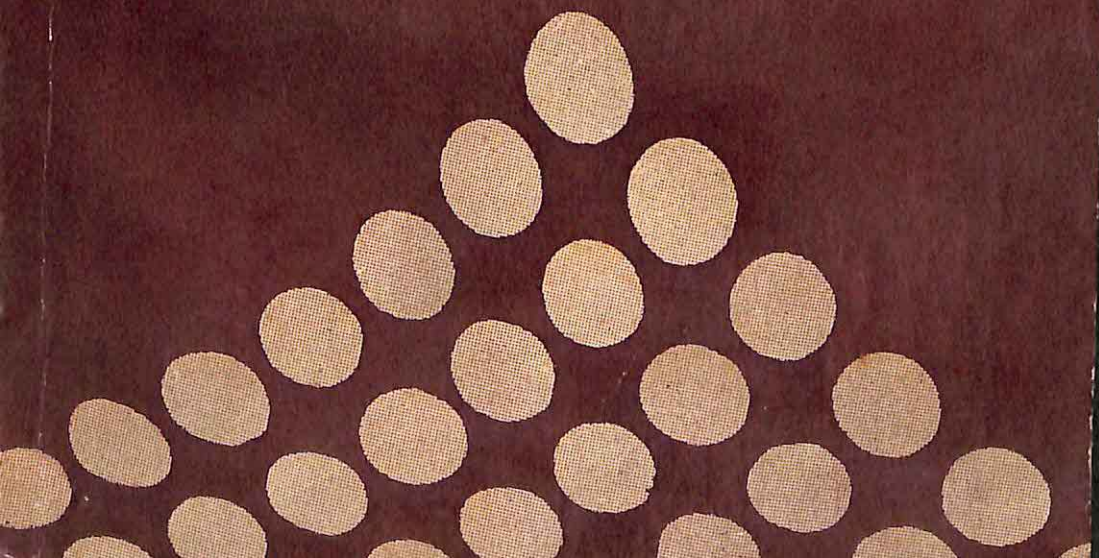
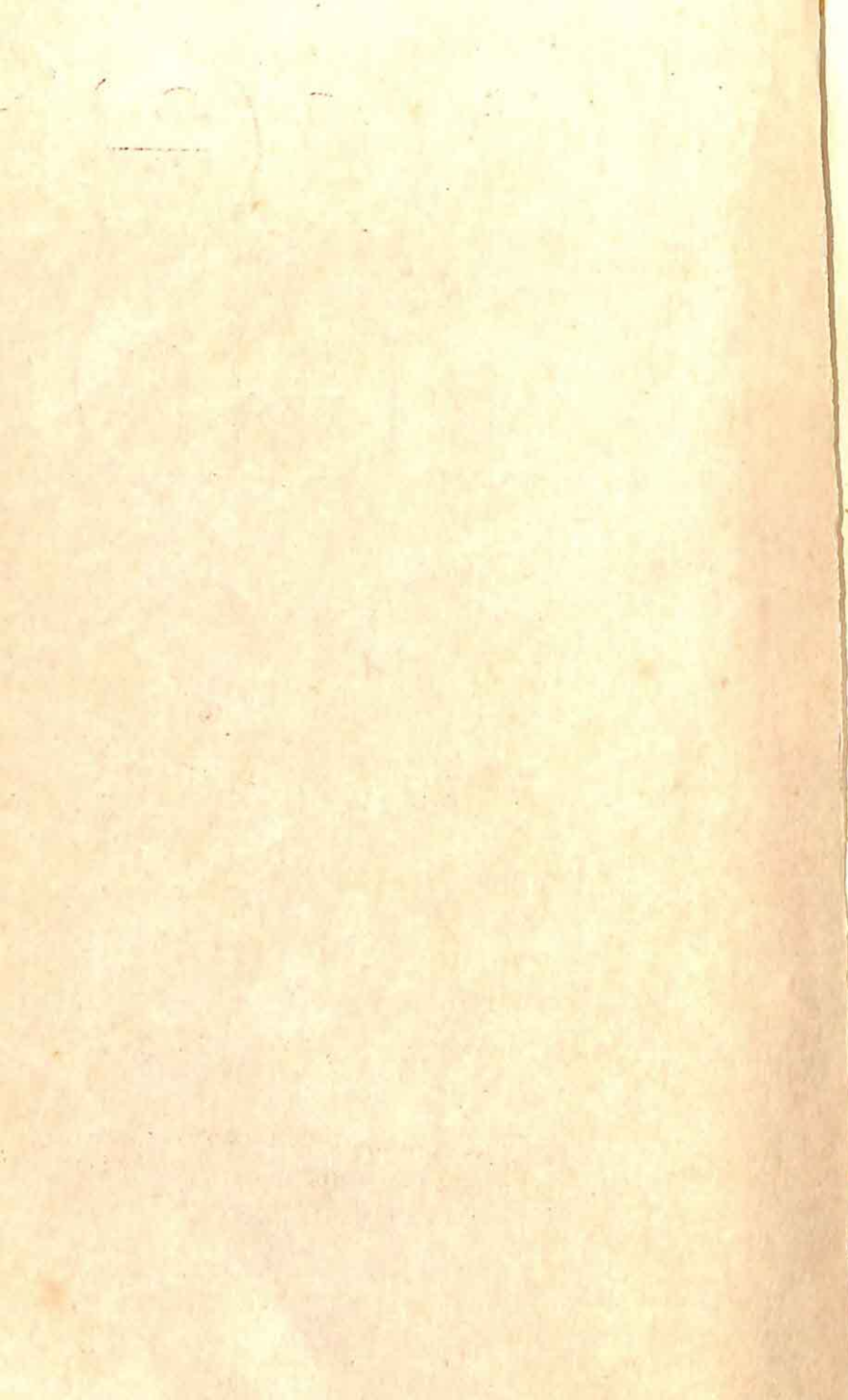


Modern Algebra

Surjeet Singh
Qazi Zameeruddin





MODERN ALGEBRA

2492



UBS Publishers' Distributors Ltd
8/1-B, Chowringhee Lane, Calcutta-700 016
Ph:212451, 249473

CHAP. IV

THE JOURNAL OF THE
SOCIETY OF THE HISTORY OF THE
CITY OF NEW YORK

Modern Algebra

(FOR HONOURS AND POSTGRADUATE STUDENTS)

SURJEET SINGH

*Professor and Head, Department of Mathematics
Guru Nanak Dev University, Amritsar*

QAZI ZAMEERUDDIN

*Department of Mathematics
K.M. College, University of Delhi*



VIKAS PUBLISHING HOUSE PVT LTD

VIKAS PUBLISHING HOUSE PVT LTD
576, Masjid Road, Jangpura, New Delhi 110014

1988 Edition

COPYRIGHT © SURJEET SINGH AND QAZI ZAMEERUDDIN, 1972

Acc. no- 16688

Printed at SONY OFFSET PRINTERS Delhi-110 032

PREFACE TO THE SIXTH REVISED EDITION

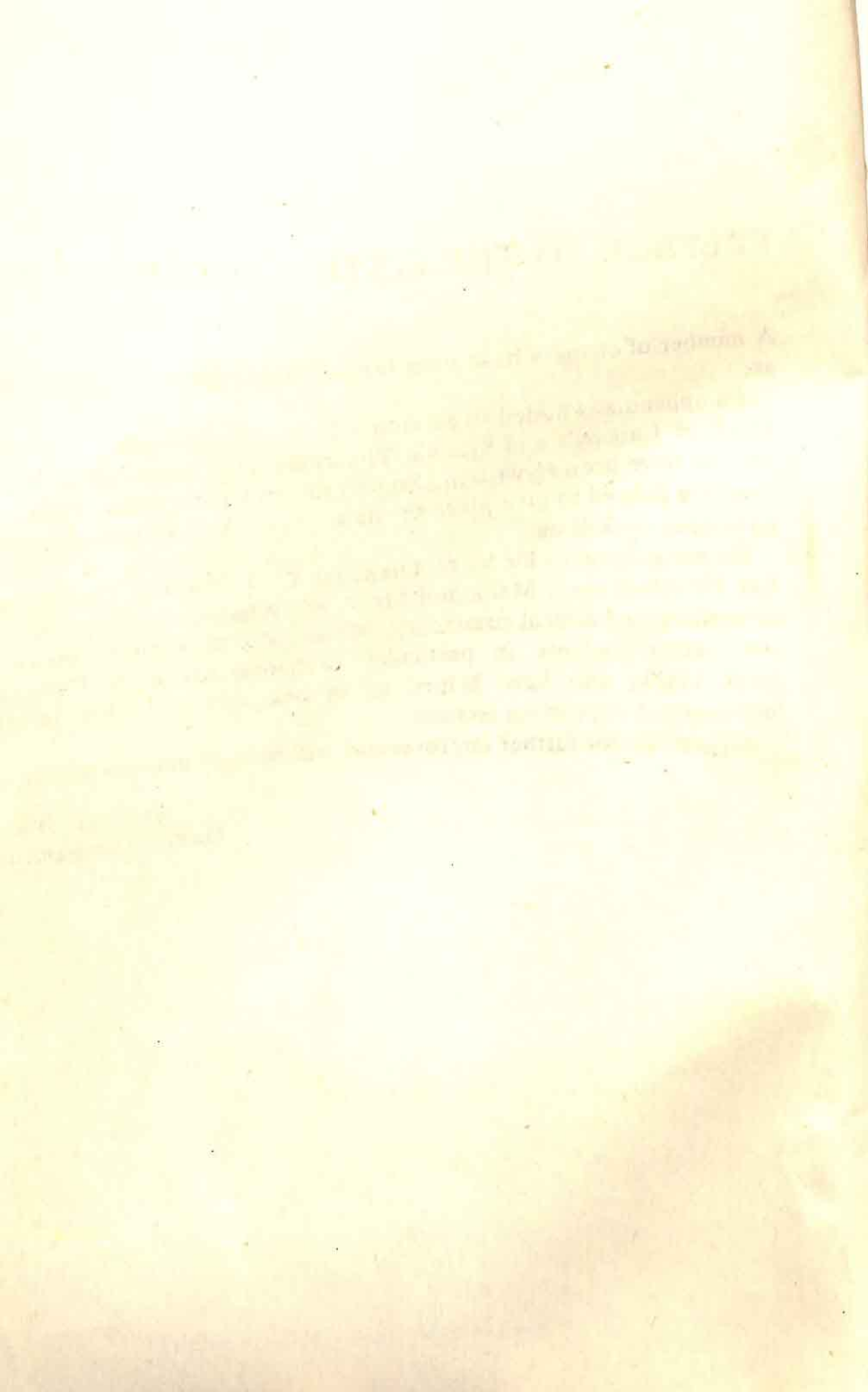
A number of changes have been made in this edition. Some of these are :

An appendix is added to Section 4.2. In this appendix alternative proofs of Cauchy's and Sylow's Theorems are provided and two sections have been re-written. Some problems have been added and some are deleted to give place to new ones. A few more exercises have been worked out.

We are grateful to Dr V. N. Dixit, Dr T. D. Narang, Dr Harbans Lal, Dr (Miss) Saroj Malik and Mr V. K. Khanna for their valuable suggestions and critical comments. We are also thankful to many of our former students, in particular to Sanjay Arora and Devender Singh Malik, who have helped us in one way or other in the improvement of problem sections.

Suggestions for further improvement will be duly acknowledged.

SURJEET SINGH
QAZI ZAMEERUDDIN



PREFACE TO THE FIRST EDITION

During the present century the developments in various aspects of Modern Algebra have been quite rapid, and so many new concepts and branches are coming up in the subject, at such a fast rate that it is almost impossible for an individual to keep track of all these ideas. In an introductory text like *ours*, it is indeed impossible even to give a glimpse of the developments which are taking place in the subject these days. All that we have discussed here can only be regarded as a classical part of Modern Algebra. However, we feel that this part is basic to all that is being developed in the subject.

In India the importance of the subject was realized only in the late fifties and in recent times large number of universities in this country have introduced courses in Modern Algebra at both undergraduate and postgraduate levels. The contents and scope of the syllabi vary significantly from place to place. It is, indeed, very difficult to say precisely as to how much an introductory course on Modern Algebra should include at undergraduate and postgraduate levels. The opinion can differ from person to person.

Whatever we have included in this textbook, we feel from our teaching experience, should be learnt by all those who want to pursue mathematics. There are many topics, like Peano's axioms on number system, some canonical forms of matrices over fields, general theory of modules, lattices, inner product spaces, etc. which can find rightful place in an introductory text on Modern Algebra. But keeping in view the scope and size of the book we felt it advisable not to include the same here.

We have made free use of elementary theory of numbers (including Euclid's division algorithm), matrices over complex numbers, calculus and Euclidean geometry in the construction of various examples and counter examples given in the book. Indeed this book can be beneficially read by those who have some knowledge of the subjects mentioned above.

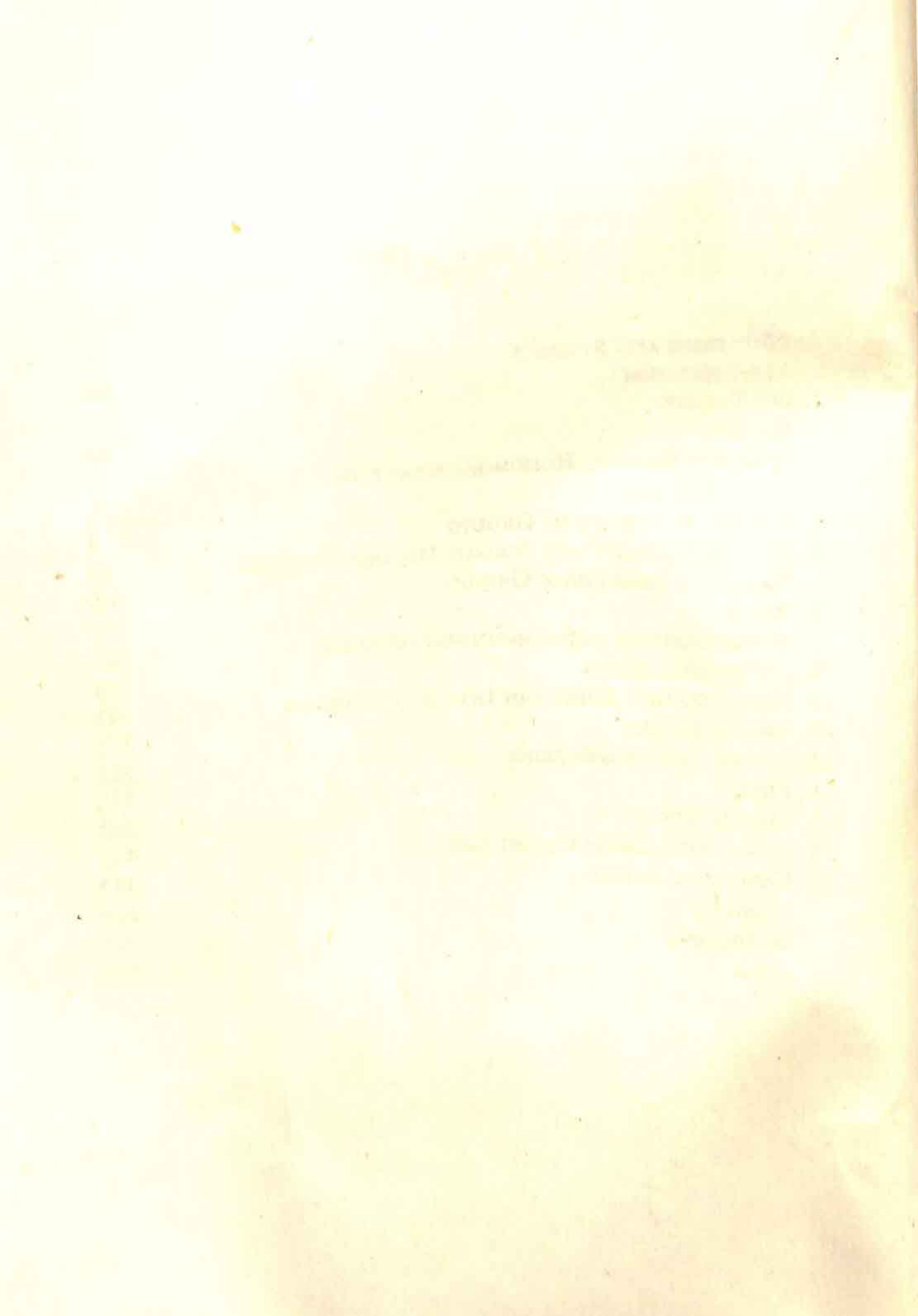
We are extremely grateful to all those mathematicians whose textbooks on Modern Algebra and research papers have influenced our learning of the subject. We are deeply thankful to our reverend teacher, Dr P.B. Bhattacharya, Department of Mathematics, Delhi University, whose constant inspiration and suggestions had a great

influence on the book. Our sincere thanks are due to all those friends and colleagues (in particular Dr Ravinder Kumar) who read some portions of the manuscript and gave their valuable suggestions. As it is our firm conviction that any form of presentation of any matter can always be improved upon, we shall gratefully accept all suggestions, comments or criticisms, which can be effectively used to improve the book.

SURJEET SINGH
QAZI ZAMEERUDDIN

CONTENTS

NOTATIONS AND SYMBOLS	<i>xi</i>
PRE-REQUISITES	<i>xii</i>
1. SET THEORY	1
2. GROUPS	42
3. QUOTIENT GROUPS, HOMOMORPHISMS AND PERMUTATIONS	75
4. STRUCTURE THEORY OF GROUPS	119
5. SOLVABLE GROUPS AND JORDAN-HOLDER THEOREM	151
6. SURVEY OF SOME FINITE GROUPS	166
7. RINGS	179
8. HOMOMORPHISMS AND EMBEDDING OF RINGS	207
9. POLYNOMIAL RINGS	233
10. FACTORIZATION THEORY IN INTEGRAL DOMAINS	243
11. VECTOR SPACES	275
12. LINEAR TRANSFORMATIONS	323
13. FIELDS	350
14. GALOIS THEORY	388
15. RINGS WITH CHAIN CONDITIONS	433
16. CANONICAL FORMS	445
<i>Appendix</i>	501
<i>Bibliography</i>	521
<i>Index</i>	525



NOTATIONS AND SYMBOLS

$\in$: Belongs to

$\forall$: For all

$\exists$: There exists

$\Rightarrow$: Implies

$\Leftrightarrow$: Implies that and is implied by

$\mathbb{Z}$: Set of all-Integers

$\mathbb{Q}$: Set of all Rational Numbers

$\mathbb{R}$: Set of all Real Numbers

$\mathbb{C}$: Set of all Complex Numbers

$\mathbb{N}$: Set of all Natural Numbers

$a \mid b$: a divides b

$\sigma \mid A$: Restriction of mapping σ to set A

ϕ : Null set

$A < B$: Set A is *strictly* contained in set B

$[E : F]$: Dimension of vector space E over a field F

$a \nmid b$: a does not divide b

■ : End of a proof.

PRE-REQUISITES

A fair knowledge of elementary properties of natural numbers, integers, rational and irrational numbers, real and complex numbers is presupposed.

Basic properties of determinants and matrices with entries from complex numbers will be frequently employed in the examples illustrating many concepts of Modern Algebra.

Acquaintance with fundamentals of real analysis will also be assumed.

Except for the topics needing the use of aforesaid concepts, the book is almost self-contained.

1

SET THEORY

1. Sets

The concept of set is most basic in Mathematics. Indeed almost all mathematical systems are certain collections of sets. All these mathematical systems and their theories can be treated as parts of set theory. In any mathematical system there are undefined terms which are related to one another by some logically consistent statements (?) called *postulates*. In the theory of sets, the concepts of *set*, *object*, *equality* and '*is an element of*' are undefined. We do not intend to discuss the axioms in set theory, as the axiomatic development of set theory is not within the scope of this book. Here we shall treat some preliminaries of set theory, which we need later in the book, rather intuitively. We do not claim any logical perfection in the proofs of various results discussed here.

Intuitively speaking a set is synonym with a *collection of objects*. Thus for us the collection of positive integers, the collection of students in a particular class, the collection of articles in a particular house, the collection of all triangles in a plane, are some examples of set. A set will be denoted by a Capital Roman letter (antique or italic) and an object or element by lower case letters.

The Symbols $\mathbf{N}$, $\mathbf{Z}$, $\mathbf{Q}$, $\mathbf{R}$, $\mathbf{C}$ will be usually reserved for sets of natural numbers, integers, rational numbers, real numbers and complex numbers respectively.

Let A be a set. If x is any object (element) which belongs to A then we write $x \in A$, and say that x is a member of A .

If x does not belong to A , then we write $x \notin A$. For example $2 \in \mathbf{N}$, but $\frac{1}{2} \notin \mathbf{N}$; however $\frac{1}{2} \in \mathbf{Q}$. If A is the set of all triangles in a plane and x is any square, then $x \notin A$. Usually the elements of a set can be specified by some property $P(x)$ in the sense that given a set A it is usually possible to specify some property $P(x)$, such that A consists of all those objects x , which satisfy $P(x)$. In that case we write

$$A = \{x \mid P(x)\}.$$

A set having no element is called a *null set* or an *empty set* and is denoted by the symbol ϕ .

EXAMPLE 1. Let $P(x) \equiv 'x \text{ is a natural number}'$. Since $\mathbf{N}$ is the set of all natural numbers, we can say that

$$\mathbf{N} = \{x \mid x \text{ is a natural number}\}.$$

EXAMPLE 2. Let A be the set of all those natural numbers which are perfect square. We can say

$$A = \{x \mid x = y^2 \text{ for some natural number } y\}.$$

If a set A consists of elements $a_1, a_2, a_3, \dots, a_n, \dots$, it is customary to write

$$A = \{a_1, a_2, a_3, \dots, a_n, \dots\}.$$

For example $A = \{1, 5, 6\}$ means A consists of 1, 5, 6 only. In this case $1 \in A$, $5 \in A$, $6 \in A$ but $2 \notin A$, $3 \notin A$, $4 \notin A$, $7 \notin A$ etc.

Definition 1.1. A set A is said to be a *subset* of a set B (symbolically $A \subseteq B$) in case every number of A is a member of B ; further if $A \neq B$, then A is said to be a *proper subset* B and we write $A \subset B$.

If A is a subset of B , we also say A is *contained in* B , or B *contains* A or B is a *superset* of A . For example $\mathbf{N} \subseteq \mathbf{Q} \subseteq \mathbf{R} \subseteq \mathbf{C}$.

It is immediate from the definition that $A \subseteq B$, $B \subseteq C$ implies $A \subseteq C$.

Two sets A and B are *equal* in case they have same elements. Thus if A and B are two sets such that $A \subseteq B$ and $B \subseteq A$, then $A \subseteq B$ gives that every member of A is a member of B and $B \subseteq A$ gives that every member of B is a member of A ; consequently A and B have same members and so $A = B$.

Definition 1.2. The *union* of two sets A and B is the set $A \cup B$, consisting of all those elements which are either in A or in B , i.e.,

$$A \cup B = \{x \mid x \in A \text{ or } x \in B\}.$$

Definition 1.3. The *intersection* of two sets A and B is the set $A \cap B$, consisting of all those x 's which are common to A and B , i.e.,

$$A \cap B = \{x \mid x \in A \text{ and } x \in B\}.$$

EXAMPLE 3. Let $A = \{1, 2, 3, 4\}$, $B = \{3, 4, 2, 5\}$.

$$A \cup B = \{1, 2, 3, 4, 5\}, \quad A \cap B = \{2, 3, 4\}.$$

EXAMPLE 4. Let A be the set of all those integers which are less than or equal to zero. Then $\mathbf{N} \cup A = \mathbf{Z}$.

Theorem 1.4. For any two sets A and B

(i) $A \subseteq B$ if and only if $A \cap B = A$.

(ii) $A \subseteq B$ if and only if $A \cup B = B$.

Proof: (i) Let $A \subseteq B$. Then for any $x \in A$, $x \in B$.

Consequently $x \in A \cap B$. This shows that $A \subseteq A \cap B$.

Clearly $A \cap B \subseteq A$. Hence $A = A \cap B$.

Conversely let $A = A \cap B$. Since in general $A \cap B \subseteq B$ so $A \subseteq B$. Similar proof holds for (ii). ■

We now establish important relationships between intersection and union, which are known as *laws of distributivity*.

Theorem 1.5. For any three sets A , B and C

$$(i) A \cap (B \cup C) = (A \cap B) \cup (A \cap C).$$

$$(ii) A \cup (B \cap C) = (A \cup B) \cap (A \cup C).$$

Proof: $x \in A \cap (B \cup C)$.

$$\Rightarrow x \in A \text{ and } x \in B \cup C$$

$$\Rightarrow x \in A \text{ and } (x \in B \text{ or } x \in C).$$

Thus if $x \in B$, then $x \in A \cap B \subseteq (A \cap B) \cup (A \cap C)$ gives $x \in (A \cap B) \cup (A \cap C)$.

Similarly if $x \in C$, since $x \in A$, we again get

$$x \in (A \cap B) \cup (A \cap C).$$

It follows that

$$A \cap (B \cup C) \subseteq (A \cap B) \cup (A \cap C). \quad \dots (1)$$

Now $x \in A \cap B$

$$\Rightarrow x \in A \text{ and } x \in B$$

$$\Rightarrow x \in A \text{ and } x \in B \cup C, \text{ since } B \subseteq B \cup C$$

$$\Rightarrow x \in A \cap (B \cup C).$$

Thus $A \cap B \subseteq A \cap (B \cup C)$ (2)

Similarly $A \cap C \subseteq A \cap (B \cup C)$ (3)

From (2) and (3) we get

$$(A \cap B) \cup (A \cap C) \subseteq A \cap (B \cup C). \quad \dots (4)$$

Then (1) and (4) yield

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C).$$

We leave it to the readers to prove (ii) on similar lines. ■

Definition 1.6. Given two sets A , B ,

the *difference set* $A - B$ is the set $\{x \in A \mid x \notin B\}$.

Further if B is a subset of A , then $A - B$ is called the *complement* of B in A .

Sometimes when there is no confusion regarding A , for any subset B of A , $A - B$ is also denoted by B' .

EXAMPLE 5. Let E be the set of all even natural numbers, then $\mathbb{N} - E$ is the set of all odd natural numbers.

EXAMPLE 6. Let $A = \{1, 2, 3, 4, 5, 6\}$

$$B = \{5, 6, 7, 8\}.$$

Then

$$A - B = \{1, 2, 3, 4\},$$

$$B - A = \{7, 8\}.$$

Clearly here $A-B$ and $B-A$ are such that they have no member in common

i.e., $(A-B) \cap (B-A) = \phi$.

Definition 1.7. Given two elements a, b , the ordered pair (a, b) with a as the first co-ordinate and b as the second co-ordinate is the set defined by

$$(a, b) = \{\{a\}, \{a, b\}\}$$

Thus (a, b) is a set of sets, whose members are the sets $\{a\}, \{a, b\}$.

For example

$$(1, 2) = \{\{1\}, \{1, 2\}\} \text{ and } (2, 1) = \{\{2\}, \{2, 1\}\}$$

Theorem 1.8. $(a, b) = (c, d)$ if and only if $a=c$ and $b=d$.

Proof: By definition,

$$\{\{a\}, \{a, b\}\} = \{\{c\}, \{c, d\}\} \quad \dots (1)$$

Case I. $a=b$.

Then

$$\begin{aligned} \{\{a\}, \{a, b\}\} &= \{\{a\}, \{a, a\}\} = \{\{a\}, \{a\}\} \\ &= \{\{a\}\}. \end{aligned} \quad \dots (2)$$

From (1) and (2) we get

$$\begin{aligned} \{\{a\}\} &= \{\{c\}, \{c, d\}\} \\ \Rightarrow \{a\} &= \{c\} = \{c, d\} \Rightarrow a=c=d \Rightarrow a=b=c=d, \text{ as } a=b. \end{aligned}$$

Thus $a=c$ and $b=d$.

Similarly if $c=d$, then $a=c$ and $b=d$.

Case II. $a \neq b$ and $c \neq d$.

Since $\{a\} \in \{\{a\}, \{a, b\}\}$ it follows from (1) that

$$\begin{aligned} \{a\} &\in \{\{c\}, \{c, d\}\} \\ \Rightarrow \{a\} &= \{c\} \text{ or } \{a\} = \{c, d\}. \end{aligned}$$

In the latter case, $a=c=d$; this is a contradiction since $c \neq d$.

So that

$$\{a\} = \{c\}$$

$$\text{and } a=c \quad \dots (3)$$

Again (1) yields $\{a, b\} \in \{\{c\}, \{c, d\}\}$

$$\Rightarrow \{a, b\} = \{c\} \text{ or } \{a, b\} = \{c, d\}.$$

In the former case $a=b=c$, a contradiction.

Hence $\{a, b\} = \{c, d\}$

$$\Rightarrow b \in \{c, d\} \Rightarrow b=c \text{ or } b=d$$

In the former case $a=b$, since $a=c$. This again is a contradiction. Consequently $b=d$.

From (3) and (4) it follows that $a=c$ and $b=d$.

Conversely if $a=c$ and $b=d$, then

$$\{a\} = \{c\} \text{ and } \{a, b\} = \{c, d\}$$

$$\Rightarrow (a, b) = \{\{a\}, \{a, b\}\} = \{\{c\}, \{c, d\}\} = (c, d).$$

Hence the result follows. ■

Purpose of the above definition and theorem on ordered pairs is to give a meaning to the order of arrangement of some elements. For example the ordered pairs $(1, 2)$ and $(2, 1)$ are different. It is well-known that in Euclidean plane the co-ordinates of a point P are given by an ordered pair (a, b) of real numbers, where a is called its x co-ordinate and b is called its y co-ordinate. The points with co-ordinates $(1, 2)$ and $(2, 1)$ are distinct. However the two sets $\{1, 2\}$ and $\{2, 1\}$ are same, since they have same members.

Definition 1.9. Given two sets A and B , the Cartesian product $A \times B$ is defined as the set of all ordered pairs of the form (a, b) where $a \in A$ and $b \in B$.

Since co-ordinates of a point in a Euclidean plane is an ordered pair (a, b) of real numbers and conversely every ordered pair of real numbers is the co-ordinates of some point, we can say that the set of all co-ordinates of all points in a Euclidean plane is $\mathbf{R} \times \mathbf{R}$.

EXAMPLE 7. Let $A = \{\alpha, \beta\}$, $B = \{1, 2, 3\}$.

Then

$$A \times B = \{(\alpha, 1), (\alpha, 2), (\alpha, 3), (\beta, 1), (\beta, 2), (\beta, 3)\}.$$

$$B \times A = \{(1, \alpha), (2, \alpha), (3, \alpha), (1, \beta), (2, \beta), (3, \beta)\}.$$

REMARK. If a set A has n elements, then we write $|A| = n$.

Thus in example 3,

$$|A| = 4, |B| = 4, |A \cup B| = 5 \text{ and } |A \cap B| = 3.$$

WORKED-OUT EXERCISES

Exercise 1. If two sets A and B are such that $|A| = m$ and $|B| = n$ then

$$|A \cup B| = |A| + |B| - |A \cap B| = m + n - |A \cap B|.$$

In particular if A and B are disjoint then $|A \cup B| = m + n$.

Solution. Note that when we count the number of elements in $A \cup B$, elements of $A \cap B$ are counted twice, once in the counting of elements of A and second time in the counting of the elements of B . So first result follows.

The second part follows from the observation that $|\phi| = 0$.

Exercise 2. A survey shows that 63% of the Americans like cheese whereas 76% like apples. What can you say about the percentage of the Americans who like both cheese and apples?

Solution. Let A denote the set of the Americans who like cheese and let B denote those who like apples. Assume that the population

of America is 100. Then $|A| = 63$ and $|B| = 76$.

$$\text{Now } |A \cup B| = |A| + |B| - |A \cap B|$$

$$\Rightarrow |A \cap B| = 139 - |A \cup B|.$$

But $|A \cup B| \leq 100$, so that $|A \cap B| \geq 39$.

Further $A \cap B \subseteq A$ and $A \cap B \subseteq B$ imply $|A \cap B| \leq |A|$ and $|A \cap B| \leq |B|$, that is

$$|A \cap B| \leq 63 \text{ and } |A \cap B| \leq 76.$$

$$\text{Hence } 39 \leq |A \cap B| \leq 63.$$

Consequently the percentage of the Americans who like both cheese and apples lie between 39 and 63.

Exercise 3. Two sets X and Y are equal if and only if

$$X \cup Y = X \cap Y.$$

Solution. In case $X=Y$, $X \cup Y=X$ and $X \cap Y=X$ and so

$$X \cup Y = X \cap Y.$$

Conversely if $X \cup Y = X \cap Y$, then $X \subseteq X \cup Y = X \cap Y \subseteq Y$ and $Y \subseteq X \cup Y = X \cap Y \subseteq X$ imply $X=Y$.

Exercise 4. For any three sets A, B, C ,

$$A - (B \cup C) = (A - B) \cap (A - C)$$

$$\text{and } A - (B \cap C) = (A - B) \cup (A - C).$$

(These are known as *De Morgan's Laws* for complements.)

Solution. $x \in A - (B \cup C) \Rightarrow x \in A$ and $x \notin (B \cup C)$

$\Rightarrow (x \in A \text{ such that } x \notin B) \text{ and } (x \in A \text{ such that } x \notin C)$

$\Rightarrow x \in A - B \text{ and } x \in A - C \Rightarrow x \in (A - B) \cap (A - C).$

Hence $A - (B \cup C) \subseteq (A - B) \cap (A - C).$

Conversely $y \in (A - B) \cap (A - C) \Rightarrow y \in A - B$ and $y \in A - C$

$\Rightarrow (y \in A \text{ but } y \notin B) \text{ and } (y \in A \text{ but } y \notin C)$

$\Rightarrow y \in A \text{ and } y \notin (B \cup C) \Rightarrow y \in A - (B \cup C).$

Hence $(A - B) \cap (A - C) \subseteq A - (B \cup C).$

Consequently $A - (B \cup C) = (A - B) \cap (A - C).$

For second part $x \in A - (B \cap C) \Rightarrow x \in A$ but $x \notin B \cap C$

$\Rightarrow$ either $(x \in A \text{ but } x \notin B)$ or $(x \in A \text{ but } x \notin C)$

$\Rightarrow x \in A - B \text{ or } x \in A - C \Rightarrow x \in (A - B) \cup (A - C)$

So we get $A - (B \cap C) \subseteq (A - B) \cup (A - C)$

Conversely $y \in (A - B) \cup (A - C) \Rightarrow$ either $y \in A - B$

or $y \in A - C \Rightarrow$ either $(y \in A \text{ and } y \notin B)$

or $(y \in A \text{ and } y \notin C) \Rightarrow y \in A \text{ and } y \notin (B \cap C)$

$\Rightarrow y \in A - (B \cap C).$

Therefore $(A - B) \cup (A - C) \subseteq A - (B \cap C).$

As a consequence we get $A - (B \cap C) = (A - B) \cup (A - C).$

Exercise 5. Let X, Y, A be three sets Such that $A \cap X = A \cap Y$ and $A \cup X = A \cup Y$ then $X=Y$.

Solution. We show that $X \subseteq Y$

$Y \subseteq X$ will follow by symmetry of given conditions.

Let $x \in X$. Two cases arise

Case I. $x \in A$.

Now $x \in X \Rightarrow x \in A \cap X = A \cap Y$
 $\Rightarrow x \in Y$.

Case II. $x \notin A$

Since $x \in X \Rightarrow x \in A \cup X = A \cup Y$
 $\Rightarrow x \in A \cup Y$ and $x' \notin A$
 $\Rightarrow x \in Y$.

Hence in each case $x \in X$ implies $x \in Y$.

Consequently $X \subseteq Y$.

PROBLEMS

- For any three sets A, B, C prove that
 - $A \cup (B \cap C) = (A \cup B) \cap C$.
 - $A \cap (B \cup C) = (A \cap B) \cup C$.
- Prove that the null set is a subset of every set.
- Given $A = \{1, 2, 3\}$, answer the following:
 - What are elements of A ?
 - How many subsets does A have and find them? (**Ans.** 8)
 - Let $B = \{3, 4, 5\}$. Is B a subset of A ? Find $A - B$ and $B - A$.
- For any n sets $A_1, A_2, \dots, A_n$, define $A_1 \cup A_2 \dots \cup A_n$ inductively as follows:

$$A_1 \cup A_2 \dots \cup A_n = (A_1 \cup A_2 \cup \dots \cup A_{n-1}) \cup A_n.$$
 Similarly define $A_1 \cap A_2 \cap \dots \cap A_n$. Prove the following:
 - $(A_1 \cup A_2) \cap A_3 = (A_1 \cap A_3) \cup A_2 = A_1 \cup A_2 \cap A_3$.
 - $(A_1 \cap A_2) \cup A_3 = (A_1 \cup A_3) \cap A_2 = A_1 \cap A_2 \cup A_3$.
- If A has n elements and B has m elements, what can you say about the largest or smallest number of elements that $A - B$ can have?
- For any two sets A and B , prove that $A - B$ and $B - A$ are disjoint in the sense that they have no member in common.
- For any three sets A, B, C prove the following:
 - $(A \cap B) \times C = (A \times C) \cap (B \times C)$.
 - $(A \cup B) \times C = (A \times C) \cup (B \times C)$.
 - $(A \times C) - (B \times C) = (A - B) \times C$.
- If A and B are two sets having m and n elements respectively, show that $A \times B$ has mn elements.

9. Given $A = \{1, 2, 5, 6\}$, $B = \{1, 2, 3\}$. Determine

$$(A \times B) \cup (B \times A), (A \times B) \cap (B \times A), (A \times B) - (B \times A).$$

10. The set of all subsets of a set A is called the *power set* of A and is denoted by $P(A)$. If A has n members show that $P(A)$ has 2^n members.

11. A, B, C are three sets such that $|A| = m$, $|B| = n$ and $|C| = p$. Show that $|A \cup B \cup C| = m + n + p - |A \cap B| - |B \cap C| - |C \cap A| + |A \cap B \cap C|$.

12. In a battle 70% of the combatants lost one eye, 80% an ear, 75% an arm, 85% a leg, $x\%$ lost all the four limbs. Find the minimum value that is possible for x . (Ans. 10)

13. In a survey of 100 students, the numbers studying the various languages were found to be: Spanish 28, German 30, French 42, Spanish and German 8, Spanish and French 10, German and French 5, all three languages 3.

(i) How many were studying no language?

(ii) How many had French as their only language?

[Ans. (i) 20; (ii) 30.]

14. Let $X \cup Y = X$ for all sets X . Prove that $Y = \phi$.

15. If P, Q, R are subsets of A and let the relative complements ($'$) be with respect to A . Prove the following:

(i) $P \subseteq Q$ if and only if $P \cap Q' = \phi$

(ii) $P \subseteq Q$ if and only if $P' \cup Q = A$

(iii) $P \subseteq Q$ if and only if $P \cap Q' \subseteq P'$

(iv) $P \subseteq Q$ if and only if $P \cap Q' \subseteq Q$

(v) $P \subseteq Q$ if and only if $P \cap Q' \subseteq R \cap R'$

16. A and B are two sets. Which of the following is true and which is false? Justify your answer.

(i) $P(A \cap B) = P(A) \cap P(B)$

(ii) $P(A \cup B) = P(A) \cup P(B)$

(iii) $P(A - B) = P(A) - P(B)$.

[Ans. (i) true (ii) false (iii) false.]

17. Let X, Y, S, T be sets such that

$$X \cup Y = T \cup S \text{ and } X \cap Y = T \cap S = \phi.$$

Prove that $X = \phi$ if and only if $T = (X \cap S) \cup (Y \cap T)$.

2. Relations and Mappings

In every day context we frequently speak of relations which hold between two or among several things. For example (i) 2 divides 4, (ii) Akbar was the father of Jehangir, (iii) 6 is greater than 4, are three examples of relation in every-day context. In these examples

we find that some sort of connections between pairs of objects are given. Thus these examples express some relations between pairs of objects. One might think of relation between several objects. This vague idea of an intuitive connectedness may be dispensed with in the formal definition of a relation. Rather we define a relation as follows:

Definition 1.10. Let X and Y be two non-empty sets. Then any subset of $X \times Y$ is called a *binary relation* of X to Y . A relation is also called a *correspondence*.

If $X=Y=S$, then a subset of $S \times S$ is called a *binary relation* on S .

In this book we are only concerned with binary relations, hence by a 'relation' we shall always mean a 'binary relation'. If T is a relation of A to B and $(a, b) \in T$, then we write aTb and say that ' a has relation T to b '. If $(a, b) \notin T$, then we say ' a does not have relation T to b '. It follows from definition that on a set S , there are as many relations as the number of subsets of $S \times S$. Among these, the equivalence relations discussed later in this section are of great importance.

Definition 1.11. Let X and Y be two non-empty sets. A subset T of $X \times Y$ is called a *mapping* of X into Y if for each $x \in X$, there exists one and only one $y \in Y$ such that $(x, y) \in T$. A mapping is also known as a *function*.

For any $x \in X$ if $(x, y) \in T$ then y is called the *value* of T at x or the *image* of x under T or y is said to *correspond* to x under T and we write $y = T(x)$. If T is a mapping of X into Y , then X is called the *domain* of T ($\text{dom } T$) and the set $\{y \in Y \mid y = T(x) \text{ for some } x \in X\}$ is called the *range* of T ($\text{range } T$). Y is called *co-domain* of T ($\text{co-dom } T$).

NOTATION. The symbol $f: X \rightarrow Y$ will denote that f is a mapping of X into Y .

Definition 1.12. Let f be a mapping of a set X into a set Y .

Then (i) f is said to be a *one-to-one* mapping if for all $x_1, x_2 \in X$, $f(x_1) = f(x_2)$ implies that $x_1 = x_2$.

(ii) If the range of f is Y then f is said to be an *onto* mapping.

(iii) If f is both one-to-one and onto then it is called a *one-to-one correspondence* (1-1 correspondence).

Theorem 1.12 (a). Two mappings f, g of a set X into a set Y are equal if and only if $f(x) = g(x)$ for all $x \in X$.

Proof: Suppose $f = g$.

Let $x \in X$, $(x, f(x)) \in f$.

Since $f = g$ we get $(x, f(x)) \in g$.

This gives $f(x)=g(x)$.

Conversely let $f(x)=g(x)$ for all $x \in X$.

Let $(a, b) \in f$. Then $b=f(a)=g(a)$ gives $(a, b) \in g$. So $f \subseteq g$.

Similarly it can be proved that $g \subseteq f$.

Hence $g=f$. ■

REMARK. A one-to-one mapping is also called an *injection* and an onto mapping is also called a *surjection*. A 1-1 correspondence is also called a *bijection*.

EXAMPLE 8. Let X be a non-void set, define $i: X \rightarrow X$ by $i(x)=x \forall x \in X$. In terms of ordered pairs

$$i = \{(x, x) \mid x \in X\}.$$

i is called the *identity* mapping on X .

EXAMPLE 9. Let $X=Y=\mathbf{Z}$. Define $f: \mathbf{Z} \rightarrow \mathbf{Z}$ by $f(x)=2x \forall x \in \mathbf{Z}$. Since for any $x_1, x_2 \in \mathbf{Z}$, $f(x_1)=f(x_2) \Rightarrow 2x_1=2x_2 \Rightarrow x_1=x_2$, f is a one-to-one mapping. Now $\text{range } f = \{f(x) \mid x \in \mathbf{Z}\} = \{2x \mid x \in \mathbf{Z}\} = \text{Set of all even integers}$. Thus $\text{range } f \neq \mathbf{Z}$ and, f is not onto.

EXAMPLE 10. Let $X=\mathbf{Z} \times \mathbf{Z}$. Define $f: \mathbf{Z} \times \mathbf{Z} \rightarrow \mathbf{Z}$ by $f(n, m)=n+m. \forall (n, m) \in \mathbf{Z} \times \mathbf{Z}$. f is onto, since for every $n \in \mathbf{Z}$, we have $f(0, n)=0+n=n$ and $n \in \text{range } f$ and so, $\text{range } f = \mathbf{Z}$. However f is not a one-to-one mapping since $f(1, 3)=1+3=4=2+2=f(2, 2)$, but $(1, 3) \neq (2, 2)$.

EXAMPLE 11. Let $X=\{x_1, x_2, x_3, x_4\}$, $Y=\{\alpha, \beta\}$. Define $f: X \rightarrow Y$ such that $f(x_1)=\alpha, f(x_2)=\beta, f(x_3)=\alpha, f(x_4)=\alpha$, f is onto but not one-to-one. Define $g: X \rightarrow Y$ by $g(x_1)=\alpha, g(x_2)=\alpha, g(x_3)=\alpha, g(x_4)=\alpha$, then $\text{range } g = \{\alpha\} \neq Y$. So g is not onto. Also g is not one-to-one.

Definition 1.13. If $f: X \rightarrow Y$ and $g: Y \rightarrow Z$ are two mappings, then their *composite* gof (also called the *product* of f and g) is a mapping of X into Z such that for every $x \in X$, $(gof)(x)=g(f(x))$.

Here in the composite gof we say that for any $x \in X$, firstly f operates on x and then g operates on $f(x)$.

Before we proceed to discuss various properties of mappings we pause to give a few examples of composite mappings.

EXAMPLE 12. Let $X=\{a, b, c\}$, $Y=\{x, y, z, u\}$, $S=\{\alpha, \beta, \gamma\}$. Define $f: X \rightarrow Y$ by $f(a)=x, f(b)=y, f(c)=z$. Define $g: Y \rightarrow S$ by $g(x)=\alpha, g(y)=\beta, g(z)=\gamma, g(u)=\alpha$. Then $gof: X \rightarrow S$ is such that $(gof)(a)=g(f(a))=g(x)=\alpha, (gof)(b)=g(f(b))=g(y)=\beta, (gof)(c)=g(f(c))=g(z)=\gamma$.

Since images of any two distinct members under gof are distinct, gof is one-to-one. Also $\text{range } gof = S$ shows that gof is onto. Thus

gof is a bijection. However neither f nor g is a bijection, g is not even one-to-one.

EXAMPLE 13. Let $X=Y=Z$. Define $f: Z \rightarrow Z$, $g: Z \rightarrow Z$ by $f(x) = -x$ and $g(x) = x+2 \forall x \in X$

Then for any $x \in X$, $(gof)(x) = g(f(x)) = g(-x) = -x+2$
and $(fog)(x) = f(g(x)) = f(x+2) = -(x+2)$.

Since $-x+2 \neq -(x+2) \forall x$, we get $fog \neq gof$.

Theorem 1.14. Let $f: U \rightarrow V$, $g: V \rightarrow X$, $h: X \rightarrow Y$ be any three mappings. Then

- $ho(gof) = (hog)$ of.
- If f and g are both one-to-one then so is gof .
- If f and g are both onto, then so is gof .
- If $i: U \rightarrow U$ and $j: V \rightarrow V$ are identity mappings then $foi = f$ and $jof = f$.

Proof: Let $u \in U$.

(a) By definition $[ho(gof)](u) = h[(gof)(u)] = h[g(f(u))]$

Also $[(hog) of](u) = (hog)(f(u)) = h[g(f(u))]$.

Thus $[ho(gof)](u) = [(hog) of](u) \forall u \in U$.

Therefore $ho(gof) = (hog) of$

(b) Let $u_1, u_2 \in U$,

$$(gof)(u_1) = (gof)(u_2) \Rightarrow g(f(u_1)) = g(f(u_2))$$

$$\Rightarrow f(u_1) = f(u_2), \text{ since } g \text{ is one-to-one}$$

$$\Rightarrow u_1 = u_2, \text{ since } f \text{ is one-to-one.}$$

Consequently gof is one-to-one.

(c) To show that gof is onto it is enough to show that every member of X is an image of some member of U under gof . Consider $x \in X$. Since g is onto, $\exists v \in V$ such that $g(v) = x$. Now as $f: U \rightarrow V$ is onto, $\exists u \in U$ such that $f(u) = v$. Then $(gof)(u) = g(f(u)) = g(v) = x$. So that gof is onto.

(d) Now for any $u \in U$, since i is identity mapping on U , we have $i(u) = u \Rightarrow f(i(u)) = f(u) \Rightarrow (foi)(u) = f(u) \Rightarrow foi = f$.

For any $u \in U$, since $f(u) \in V$ and j is identity map on V , we have $j(f(u)) = f(u)$, i.e., $(jof)(u) = f(u)$. Hence $jof = f$.

This proves the theorem. ■

Thus in the above theorem if both f and g are bijections then (b) and (c) give gof is also a bijection. So we get.

Corollary 1.15. If $f: U \rightarrow V$ and $g: V \rightarrow X$ are both bijections then $gof: U \rightarrow X$ is also a bijection. ■

Definition 1.16. Given any set X , any one-to-one mapping of X onto X itself is called a *non-singular transformation or permutation* of X .

The symbol $S(X)$ will denote the set of all permutations of X . We have the following.

Theorem 1.17. Let X be a non-void set and let $f, g, h \in S(X)$. Then

(a) $f \circ g \in S(X)$.

(b) $(f \circ g) \circ h = f \circ (g \circ h)$.

(c) If $i : X \rightarrow X$ is the identity map on X , then $i \circ f = f = f \circ i$.

There exists $f' \in S(X)$ such that $f \circ f' = f' \circ f = i$.

Proof: (a) Since f and g are one-to-one as well onto, by parts (b) and (c) of Theorem 1.14 we get $g \circ f$ is one-to-one as well onto. So $g \circ f \in S(X)$.

(b) follows from part (a) of Theorem 1.14.

(c) follows from part (d) of Theorem 1.14.

(d) Define $f' : X \rightarrow X$ as follows. Consider $x \in X$. Since $f : X \rightarrow X$ is onto $\exists y \in X$ such that $f(y) = x$. As f is 1-1, y is unique.

Now define $f'(x) = y$,

i.e., for any $x, y \in X$, define $f'(x) = y$ if and only if $f(y) = x$.

Consider $x_1, x_2 \in X$. Let $f'(x_1) = y_1, f'(x_2) = y_2$. So $f(y_1) = x_1$ and $f(y_2) = x_2$.

$f'(x_1) = f'(x_2) \Rightarrow y_1 = y_2 \Rightarrow f(y_1) = f(y_2) \Rightarrow x_1 = x_2$. So that f' is one-to-one. Since given $x \in X$, if $f(x) = y$, then by definition $f'(y) = x$; we get $x \in \text{range } f'$; thus $X \subseteq \text{range } f'$ and $X = \text{range } f'$. Hence f' is a permutation of X and $f' \in S(X)$. Now for any $x, y \in X$, $f(x) = y$ if and only if $f'(y) = x$ gives $(f' \circ f)(x) = f'(f(x)) = f'(y) = x = i(x)$ and $(f \circ f')(y) = f(f'(y)) = f(x) = y = i(y)$

Hence $f \circ f' = f' \circ f = i$

This proves the theorem. ■

Definition 1.18. Let $f : X \rightarrow Y$ be a mapping. For any subset A of X , the set $\{f(a) \mid a \in A\}$ which is a subset of Y , is called the *image* of A under f and it is denoted by $f(A)$. For any subset B of Y , the set $\{x \in X \mid f(x) \in B\}$ which is a subset of X , is called the *pre-image* of B and it is denoted by $f^{-1}(B)$.

We state the following results without proof. Proof is left to the readers.

Theorem 1.19. Consider a mapping $f : X \rightarrow Y$. For any subsets A_1, A_2 of X, B_1, B_2 of Y

(a) $f(\phi) = \phi$.

(b) $A_1 \subseteq A_2$ implies $f(A_1) \subseteq f(A_2)$.

(c) $f(A_1 \cup A_2) = f(A_1) \cup f(A_2)$.

(d) $f(A_1 \cap A_2) \subseteq f(A_1) \cap f(A_2)$.

(e) $f^{-1}(Y) = X$.

$$(f) f^{-1}(B_1 \cup B_2) = f^{-1}(B_1) \cup f^{-1}(B_2).$$

$$(g) f^{-1}(B_1 \cap B_2) = f^{-1}(B_1) \cap f^{-1}(B_2).$$

$$(h) B_1 \subseteq B_2 \text{ implies } f^{-1}(B_1) \subseteq f^{-1}(B_2). \blacksquare$$

EXAMPLE 14. Let $X = \{1, 2, 3\}$, $Y = \{\alpha, \beta\}$.

Define $f: X \rightarrow Y$ such that $f(1) = \alpha$, $f(2) = \alpha$, $f(3) = \beta$.

Consider $A_1 = \{1\}$, $A_2 = \{2\}$. Then $f(A_1) = \{f(1)\} = \{\alpha\} = f(A_2)$. So $f(A_1) \cap f(A_2) = \{\alpha\}$. However $A_1 \cap A_2 = \emptyset$ gives $f(A_1 \cap A_2) = f(\emptyset) = \emptyset$.

Thus $f(A_1 \cap A_2) \neq f(A_1) \cap f(A_2)$.

Let $B_1 = \{\alpha\}$, $B_2 = Y = \{\alpha, \beta\}$.

Since $f(1) = \alpha = f(2)$, we get

$f^{-1}(B_1) = \{1, 2\}$. However $f^{-1}(B_2) = \{1, 2, 3\}$, so we have $f^{-1}(B_1) \subseteq f^{-1}(B_2)$ but $f^{-1}(B_1) \neq f^{-1}(B_2)$.

Theorem 1.19 (a). Let $f: X \rightarrow Y$ be a 1-1 onto function. Then there exists a function $g: Y \rightarrow X$ such that for any $y \in Y$, $g(y) = x$, $x \in X$, whenever $f(x) = y$.

Proof: Since f is onto, for any $y \in Y$, there exists $x \in X$ such that $y = f(x)$. This by definition, implies that $g(y) = x \in X$. Hence g is a relation from X to Y .

To prove that g is a function we suppose that y_1 and $y_2 \in Y$ are such that $y_1 = y_2$, $g(y_1) = x_1$ and $g(y_2) = x_2$ for some $x_1, x_2 \in X$.

Again the definition of g gives that $f(x_1) = y_1 = y_2 = f(x_2)$. As f is 1-1, we at once get $x_1 = x_2$. Hence g is a function from Y to X . $\blacksquare$

Definition 1.19(b). A 1-1 onto function $f: X \rightarrow Y$ is called *invertible function*.

REMARK. By Theorem 1.19 (a), there exists a function $g: Y \rightarrow X$ defined by $g(y) = x$ whenever $f(x) = y$; $x \in X$, $y \in Y$. This g is called an inverse of f . In next theorem we show that inverse of an invertible function is *unique* and further it is also 1-1 and onto.

Theorem 1.19 (c). For any invertible function $f: X \rightarrow Y$, inverse is *unique*.

Proof: Let g and h be two inverses of $f: X \rightarrow Y$. By definition for any $y \in Y$, $g(y) = x$, $x \in X$ whenever $f(x) = y$ and $h(y) = x_1$, $x_1 \in X$ whenever $f(x_1) = y$. Thus $f(x) = f(x_1)$. As f is 1-1, we get $x = x_1$, hence $g(y) = h(y)$.

Consequently $g = h$. $\blacksquare$

NOTATION. If f is an invertible function then its inverse is denoted by f^{-1} .

Following theorem provides an equivalent definition of an invertible function.

Theorem 1.19(d). A function $f: X \rightarrow Y$ is invertible if and only if there exists a function $g: Y \rightarrow X$ such that $gof = i_x$ and $fog = i_y$ where i_x is identity function on X and i_y is identity function on Y .

Proof: Suppose $f: X \rightarrow Y$ is invertible. By Definition 1.19(b) there exists $g: Y \rightarrow X$ such that $g(y) = x$ whenever $f(x) = y$; $x \in X$, $y \in Y$. For any $x \in X$, $gof(x) = g[f(x)] = g(y) = x$ and for any $y \in Y$ $fog(y) = f[g(y)] = f(x) = y$

Hence $gof = i_x$ and $fog = i_y$.

Conversely let $g: Y \rightarrow X$ be a function such that

$$gof = i_x \text{ and } fog = i_y.$$

Let $x_1, x_2 \in X$ be such that $f(x_1) = f(x_2)$

$$\begin{aligned} \text{Then } g[f(x_1)] &= g[f(x_2)] \Rightarrow (gof)(x_1) = (gof)(x_2) \\ &\Rightarrow i_x(x_1) = i_x(x_2) \Rightarrow x_1 = x_2 \end{aligned}$$

Hence f is 1-1.

Finally let $y \in Y$, then $y = i_y(y)$ gives $y = [fog](y) = f[g(y)]$.

As $g(y) \in X$, f is onto.

Consequently f is invertible by definition 1.19(b). ■

Sometimes the range of a function is deemed to be more important than the function itself; in that case we change the terminology and use a new term called *family* which we define as follows:

Definition 1.20. Let I and X be two sets. Then a function $f: I \rightarrow X$ is called a *family*, each member of I is called an *index*. I is called an *index set*, the range of f is called an *indexed set* (indexed by I). For each $i \in I$, $f(i)$ is called *ith term* of the family f .

If $f: I \rightarrow X$ is a family, then, for each $i \in I$, it is customary to denote $f(i)$ by f_i and we say that $\{f_i\}_{i \in I}$ is a family. If the range of f is a set of sets then f is called a *family of sets*. Thus whenever we say that $\{S_\alpha\}_{\alpha \in \Lambda}$ is a family of sets indexed by the set Λ , this means there is a function f from Λ into a set X of sets such that $f(\alpha) = S_\alpha$ for all $\alpha \in \Lambda$.

Definition 1.21. Let $\{S_\alpha\}_{\alpha \in \Lambda}$ be a family of sets then the union of this family is the set $\bigcup_{\alpha \in \Lambda} S_\alpha$ consisting of all those elements x which belong to at least one S_α . The intersection of this family, in the set $\bigcap_{\alpha \in \Lambda} S_\alpha$ consisting of all those elements x which are common to all S_α .

EXAMPLE 15. Let $A_1 = \{\alpha, \beta\}$, $A_2 = \{\alpha, \beta, \gamma\}$, $A_3 = \{\alpha, \gamma\}$, $A_4 = \{\alpha, \beta, \delta\}$, $I = \{1, 2, 3, 4\}$, $X = \{\alpha, \beta, \gamma, \delta\}$.

Define $f: I \rightarrow P(X)$ by

$$f(1) = A_1, f(2) = A_2, f(3) = A_3, f(4) = A_4.$$

Then we can say that f is a family of sets indexed by I . We can

write $f = \{A_i\}_{i \in I}$

The union of this family is

$$\{\alpha, \beta, \gamma, \delta\} = A_1 \cup A_2 \cup A_3 \cup A_4$$

and intersection of this family is $\{z\}$.

Again we turn our attention to general relations.

Definition 1.22. Let R be a relation on a set X . Then

- (a) R is said to be *reflexive*, if $(x, x) \in R$ for every $x \in X$.
- (b) R is said to be *symmetric*, if for every $x, y \in X$, $(x, y) \in R$ implies $(y, x) \in R$.
- (c) R is said to be *transitive* if $(x, y) \in R$, $(y, z) \in R$ imply $(x, z) \in R$ for $x, y, z \in X$.
- (d) R is said to be *anti-symmetric* if for all $x, y \in X$, $(x, y) \in R$ and $(y, x) \in R$ imply $x = y$.

Definition 1.23. A relation R on a set X is said to be an *equivalence relation* if it is reflexive, symmetric and transitive.

Definition 1.24. A relation R on a set X is said to be a *partial ordering* if R is reflexive, anti-symmetric and transitive.

NOTATION. A partial ordering is usually denoted by $\leq$.

If $a \leq b$ but $a \neq b$ then we write $a < b$.

EXAMPLE 16. The ordinary relation of equality of two elements in a set S is an equivalence relation on S .

EXAMPLE 17. Let X be the set of all lines in a plane. For any two lines a, b define $a \sim b$ if and only if a is parallel to b . Then for any three $a, b, c \in X$, the following holds:

- (i) Since a is parallel to a itself, $a \sim a$.
- (ii) $a \sim b \Rightarrow a$ is parallel to $b \Rightarrow b$ is parallel to $a \Rightarrow b \sim a$.
- (iii) $a \sim b, b \sim c \Rightarrow a$ is parallel to b and b is parallel to $c \Rightarrow a$ is parallel to $c \Rightarrow a \sim c$.

Thus $\sim$ is reflexive, symmetric as well as transitive. Hence $\sim$ is an equivalence relation on X .

EXAMPLE 18. Let n be a fixed positive integer. For any two integers a, b , define $a \equiv b \pmod{n}$ if and only if n divides $a - b$. It can be immediately verified that for any three integers a, b, c ,

- (i) $a \equiv a \pmod{n}$
- (ii) $a \equiv b \pmod{n} \Rightarrow b \equiv a \pmod{n}$
- (iii) $a \equiv b \pmod{n}, b \equiv c \pmod{n} \Rightarrow a \equiv c \pmod{n}$.

Thus the relation $\equiv$ on $\mathbb{Z}$ is an equivalence relation. This relation is known as the *relation of congruence modulo n* .

EXAMPLE 19. For any $a, b \in \mathbb{N}$, the set of natural numbers define $a \leq b$ if and only if a divides b . Then for any three natural

numbers a, b, c

(i) Since $a=1a$ implies a divides a , we get $a \leq a$.

(ii) $a \leq b, b \leq a \Rightarrow a$ divides b and b divides a .
 $\Rightarrow a=b$.

(iii) $a \leq b, b \leq c \Rightarrow a$ divides b and b divides c
 $\Rightarrow a$ divides $c \Rightarrow a \leq c$.

Thus ' $\leq$ ' defined above is reflexive, anti-symmetric and transitive. on N , hence ' $\leq$ ' is a partial ordering on N . Here notice that, as neither 2 divides 3 nor 3 divides 2; by definition $2 \nless 3$ and $3 \nless 2$.

Definition 1.25. Let $\sim$ be an equivalence relation on a non-empty set X . For any $a \in X$, the *equivalence class* of a under the relation $\sim$ is the set $a / \sim = \{b \in X \mid a \sim b\}$.

When there is no ambiguity concerning the equivalence relation $\sim$ on X , we shall denote the equivalence class $a / \sim$ by $Cl(a)$.

Theorem 1.26. Let $\sim$ be an equivalence relation on a set X . For any $a, b \in X$

(i) $a \in Cl(a)$.

(ii) Either $Cl(a) \cap Cl(b) = \phi$ or $Cl(a) = Cl(b)$ (i.e., any two equivalence classes are either disjoint or equal).

(iii) X is the union of the equivalence classes $Cl(a)$.

Proof: (i) By definition $Cl(a) = \{b \in X \mid a \sim b\}$. Since $a \sim a$, by reflexivity, we have $a \in Cl(a)$.

(ii) $Cl(a) \cap Cl(b) \neq \phi \Rightarrow \exists x \in Cl(a) \cap Cl(b)$
 $\Rightarrow a \sim x$ and $b \sim x$
 $\Rightarrow a \sim x$ and $x \sim b$ (by symmetry)
 $\Rightarrow a \sim b$. (by transitivity)

Let $y \in Cl(b)$. Then $b \sim y$. Using (1) and the transitivity of $\sim$, we get $a \sim y$, i.e., $y \in Cl(a)$ (1)

Thus $Cl(b) \subseteq Cl(a)$.

Similarly, interchanging the role of a and b we get $Cl(a) \subseteq Cl(b)$. Hence $Cl(a) = Cl(b)$.

(iii) Let $Y = \bigcup_{a \in X} Cl(a)$. Clearly $Y \subseteq X$.

Since for each $a \in X$, $a \in Cl(a) \Rightarrow a \in Y$ we get $X \subseteq Y$. Hence $X=Y$.

This proves the theorem. ■

In the above theorem we have seen that an equivalence relation on a set X decomposes X into disjoint equivalence classes. In general given a set X , a set F of non-empty subsets of X is called a *partition* of X , in case every distinct pair of members of F are disjoint and X is the union of members of F . Thus if we take $F = \{Cl(a) \mid a \in X\}$,

in the theorem above, we see that F is a partition of X . Hence each equivalence relation on a set X determines a partition of X . We prove the converse.

Theorem 1.27. Let $F = \{S_\alpha\}_{\alpha \in \Lambda}$ be a partition of X . Then there exists an equivalence relation on X , such that F is the set of all equivalence classes under that relation.

Proof: For any $a, b \in X$, define $a \sim b$ if and only if a, b are in same S_α . Then for any a, b, c in X , we have:

(i) *Reflexivity.* Since by definition $X = \bigcup S_\alpha$, $\exists \alpha \in \Lambda$ such that $a \in S_\alpha$. So that by definition, $a \sim a$.

(ii) *Symmetry.* $a \sim b \Rightarrow \exists \alpha \in \Lambda$ such that $a, b \in S_\alpha \Rightarrow b, a \in S_\alpha \Rightarrow b \sim a$.

(iii) *Transitivity.* $a \sim b, b \sim c \Rightarrow \exists \alpha, \beta \in \Lambda$ such that $a, b \in S_\alpha$ and $b, c \in S_\beta$. If $\alpha \neq \beta$ then by definition $S_\alpha \cap S_\beta = \phi$.

However $b \in S_\alpha \cap S_\beta$; so $\alpha = \beta$. Consequently $a, c \in S_\alpha$ and $a \sim c$.

Hence $\sim$ is an equivalence relation on X .

Consider any S_α and $a \in S_\alpha$. Now $Cl(a) = \{b \in X \mid a \sim b\}$.

Since $a \sim b$ if and only a and b are both in the same member of F , as $a \in S_\alpha$ we get $b \in S_\alpha$. Hence $Cl(a) = S_\alpha$. So that each S_α is an equivalence class. Conversely since given any $b \in X$, $b \in S_\beta$ for some $\beta \in \Lambda$ we have $Cl(b) = S_\beta$. Hence F is the set of all equivalence classes under $\sim$. This completes the proof. ■

Definition 1.27 (a). Two sets X and Y are called *Equipotent* if there exists a 1-1 correspondence between X and Y .

NOTATION: If X and Y are equipotent, we write $X \sim Y$.

EXAMPLE 19 (a). $\mathbf{Z}$ is equipotent to E , the set of even integers, since the function $f: \mathbf{Z} \rightarrow E$ given by $f(x) = 2x$ is a 1-1 correspondence.

Definition 1.27 (b). A set X is said to be *infinite* if it is equipotent to a proper subset of itself. A set X is said to be *finite* if it is equipotent to a set $\{1, 2, 3, \dots, n\}$ for any natural number n .

REMARK: Empty set is regarded as a finite set.

EXAMPLE 19 (b). $\mathbf{Z}$ is an infinite set as E is a proper subset of $\mathbf{Z}$ and $f: \mathbf{Z} \rightarrow E$ given by $f(x) = 2x$ for all $x \in \mathbf{Z}$, is a 1-1 correspondence. The set $\{a, b, c, d\}$ is a finite set as the mapping $g: \{a, b, c, d\} \rightarrow \{1, 2, 3, 4\}$ given by $g(a) = 2, g(b) = 1, g(c) = 3, g(d) = 4$ is a 1-1 correspondence.

Definition 1.27 (c). A set X is called *denumerable* if it is equipotent to $\mathbf{N}$, the set of natural numbers.

Definition 1.27 (d). A set X is called *countable* if it is equipotent to a subset of $\mathbf{N}$. Thus every denumerable set is countable

while a countable set need not be denumerable (give an example !)

Definition 1.27 (e). If $f: X \rightarrow Y$ is a function and T is a subset of X , then a function $g: T \rightarrow Y$ is called *Restriction* of f to T if $g(t)=f(t)$ for all $t \in T$. We denote g by f_T .

WORKED-OUT EXERCISES

Exercise 1. Let A and B be two sets each having n elements. Show that the number of bijections from A to B is $n!$

Solution. We apply induction on n . For $n=1$, $A=\{a_1\}$ and $B=\{b_1\}$. There is only one bijection $f: A \rightarrow B$ given by $f(a_1)=b_1$.

Let $n > 1$ and let the theorem be true for all sets having $n-1$ elements. Let $A=\{a_1, a_2, \dots, a_n\}$ and $B=\{b_1, b_2, \dots, b_n\}$. Let f be a bijection from A to B . Suppose that $f(a_1)=b_i$ for some i , $1 \leq i \leq n$. Since f is bijection from A to B , f_T is also a bijection from the subset $Y=A-\{a_1\}$ to $B-\{b_i\}$. Thus the number of bijections from A to B which take a_1 to b_i is equal to the number of bijections of $A-\{a_1\}$ to $B-\{b_i\}$. By induction hypothesis, therefore, number of bijections of A to B which take a_1 to b_i is equal to $(n-1)!$ As i can take up any value from 1 to n and a_1 can go to any of $b_1, b_2, \dots, b_n$; the total number of bijections from A to B is $n(n-1)! = n!$

Exercise 2. If X is an infinite set then every superset of X is also an infinite set.

Solution. Since X is infinite set, there exists a proper subset Y of X such that $f: X \rightarrow Y$ is a 1-1 correspondence.

Let T be a superset of X .

Define $g: T \rightarrow Y \cup (T-X)$ by $g(t)=f(t)$ for $t \in X$
 $=t$ for $t \in T-X$

Clearly g is 1-1 and onto.

Since Y is a proper subset of X , there is an element $y \in X$ such that $y \notin Y$.

Now $y \in T$ but $y \notin T-X$.

Hence $Y \cup (T-X)$ is a proper subset of T .

As a consequence T is an infinite set.

Exercise 3. If B is a subset of A and $f: A \rightarrow B$ is a 1-1 function then there exists a 1-1 correspondence between A and B .

Solution. $A=B$ then i_A is the required correspondence.

Now let $A \neq B$.

We define $f^0 = i_A$ and for any positive integer $k \geq 1$, $f^k = f^{k-1}(f)$.

Put $C = \{a \in A \mid \text{there exists a non-negative integer } n \text{ such that } a = f^n(x) \text{ for some } x \in A-B\}$.

Now for each $y \in A$, define $g(y) = f(y)$ if $y \in C$

$$=y \quad \text{if } y \notin C$$

Evidently $A-B \subseteq C$.

For each z in C there is only one non-negative integer n and only one element x in $A-B$ such that $z=f^n(x)$. Let $t \in f(C)$ then $t=f(c)$ for some $c \in C$, so $c=f^k(x)$ for some integer $k \geq 0$, $x \in A-B$. Thus $t=f^{k+1}(x) \Rightarrow t \in C$. So $f(C) \subseteq C$. Now for each $u \in A$ either $u \in B$ or $u \in A-B$. In case $u \in A-B$, $u \in C$, $g(u)=f(u) \in B$. In case $u \in B$ either $u \in C$ or $u \notin C$. In each case by definition of g , $g(u) \in B$. By construction g is 1-1. Again let $b \in B$. In case $b \notin C$, $b=g(b)$, and in case $b \in C$, $b=f^s(x)$ for some $x \in A-B$. s cannot be zero otherwise $b=x \in A-B$ which is absurd. So $s \geq 1$, $b=f[f^{s-1}(x)]=g[f^{s-1}(x)]$. Hence g is a 1-1 correspondence.

Exercise 4. (Schröder-Bernstein's Theorem) If A and B are two sets such that $A \sim X$ for some $X \subseteq B$ and $B \sim Y$ for some $Y \subseteq A$, then $A \sim B$.

Solution. Let $f: A \rightarrow X$ be a 1-1 correspondence and $g: B \rightarrow Y$ be a 1-1 correspondence. Now $gf: A \rightarrow Y$ is a 1-1 function as $X \subseteq B$. Since $Y \subseteq A$ and $gf: A \rightarrow Y$ is 1-1, we get by exercise 3, a 1-1 correspondence $h: A \rightarrow Y$. Then $g^{-1}h: A \rightarrow B$ is a 1-1 and onto function. Hence $A \sim B$.

Exercise 5. The set of all real numbers x , $0 < x < 1$ is not denumerable.

Solution. Suppose $X=\{x \mid 0 < x < 1\}$ is denumerable. So $X \sim \mathbf{N}$. Hence $\mathbf{N} \sim X$. Let $f: \mathbf{N} \rightarrow X$ be a 1-1 correspondence.

$$\text{Let } f(1) = \cdot C_{11} C_{12} C_{13} \dots$$

$$f(2) = \cdot C_{21} C_{22} C_{23} \dots$$

$$f(3) = \cdot C_{31} C_{32} C_{33} \dots$$

.....
In general $f(k) = \cdot C_{k1} C_{k2} C_{k3} \dots$ for any integer $k \geq 1$, where C_{ji} denotes the i th decimal digit of $f(j)$. Consider a real number $\cdot d_1 d_2 d_3 \dots$ where for all $i=1, 2, 3, \dots$, $d_i \neq C_{ii}$, $d_i \neq 0$ and $d_i \neq 9$. Since f is onto, there exists an integer $n \geq 1$ such that $\cdot d_1 d_2 d_3 \dots = f(n) = \cdot C_{n1} C_{n2} C_{n3} \dots$. However, if two decimal expressions are representations of the same number, then they must be identical expansions, unless in one of the expansions the digit 0 is repeated from some point on and in the other the digit 9; e.g. $\cdot 120000 \dots = \cdot 11999 \dots$. Since this case is excluded by the assumption that $d_i \neq 0$ and $d_i \neq 9$, we therefore have $d_i = C_{ni}$ for all $i=1, 2, \dots$ and in particular $d_n = C_{nn}$ which contradicts $d_i \neq C_{ii}$. Hence f is not onto. Consequently X is not denumerable.

Exercise 6. $\mathbf{R} \sim X$ where $X=\{x \mid 0 < x < 1\}$

Solution. Define $f: X \rightarrow \mathbf{R}$, by

$$f(x) = \frac{2x-1}{1-|2x-1|},$$

where $|2x-1| = 2x-1$ for $x \geq \frac{1}{2}$ and
 $= 1-2x$ for $x \leq \frac{1}{2}$.

In the first case $1-|2x-1| = 2-2x = 2(1-x) > 0$ as $0 < x < 1$.
 In the second case $1-|2x-1| = 2x > 0$. In either case denominator is positive.

Let $f(x)=f(y)$. Then $\frac{2x-1}{1-|2x-1|} = \frac{2y-1}{1-|2y-1|}$
 gives that both $2x-1$ and $2y-1$ must be positive or negative simultaneously. Let $x, y \geq \frac{1}{2}$.

$$\begin{aligned} f(x)=f(y) &\Rightarrow \frac{2x-1}{2-2x} = \frac{2y-1}{2-2y} \\ &\Rightarrow (2x-1)(1-y) = (2y-1)(1-x) \\ &\Rightarrow x=y. \end{aligned}$$

$$\begin{aligned} \text{In case } x, y \leq \frac{1}{2}; f(x)=f(y) &\Rightarrow \frac{2x-1}{2x} = \frac{2y-1}{2y} \\ &\Rightarrow x=y. \text{ Hence } f \text{ is } 1-1. \end{aligned}$$

To prove f is onto, we proceed as under:

Let $y \in \mathbf{R}$. If $y \leq 0$, then $\frac{1}{2(1-y)} \leq \frac{1}{2}$.

$$\text{and } f\left[\frac{1}{2(1-y)}\right] = y$$

Further if $y \geq 0$ then $\frac{2y+1}{2(1+y)} \geq \frac{1}{2}$

$$\text{and } f\left[\frac{2y+1}{2(1+y)}\right] = \frac{\frac{2y+1}{1+y} - 1}{2 - \frac{2y+1}{1+y}} = \frac{y}{(1+y)} = y$$

Hence $X \sim \mathbf{R}$. Consequently $\mathbf{R} \sim X$.

Exercise 7. $\mathbf{R}$ is uncountable.

Solution. If $\mathbf{R}$ were countable, $\mathbf{R}$ would have been equipotent to some subset Y of $\mathbf{N}$. If $Y=\mathbf{N}$ then by exercise 6, $X \sim \mathbf{N}$ which would be against the result of exercise 5.

Let $Y \subset \mathbf{N}$, then as $\mathbf{N} \sim \mathbf{N} \subseteq \mathbf{R}$, by Schroder-Bernstein theorem we shall get $\mathbf{R} \sim \mathbf{N}$ which is again absurd in view of exercises 5 and 6.

PROBLEMS

1. Let S and T be two sets having m and n elements respectively. How many relations of S into T exist?
 (Ans. 2^{mn})

2. Let S and T have m and n elements respectively, show that there exist n^m mappings of S into T .

3. $f: X \rightarrow Y$ is an onto mapping and X is finite, show that Y cannot have more members than X .

4. Consider the sets $X = \{1, 2, 3, 4\}$, $Y = \{\alpha, \beta, \gamma\}$.

Let $f_1 = \{(1, \alpha), (2, \alpha), (3, \beta), (4, \gamma)\}$

$f_2 = \{(1, \alpha), (1, \beta), (2, \alpha), (3, \beta), (4, \beta)\}$

$f_3 = \{(1, \beta), (2, \beta), (3, \beta), (4, \gamma)\}$

$f_4 = \{(1, \alpha), (2, \beta), (3, \gamma)\}$.

Answer the following:

(i) Which of these f_i 's are mappings of X into Y ? What are their ranges?

(ii) Which of the mappings are onto?

(iii) If some f_i is not a mapping, explain why it is not so?

5. Let $f: X \rightarrow Y$ be a mapping. On X define $x_1 \sim x_2$ if and only if $f(x_1) = f(x_2)$. Show that $\sim$ is an equivalence relation on X .

6. Let $f: X \rightarrow Y$, $g: Y \rightarrow Z$ be two functions such that $g \circ f$ is 1-1 and onto. Show that f is 1-1 and g is onto. Give examples to show that f need not be onto and g need not be 1-1 while $g \circ f$ is 1-1 and onto.

7. Let X be the set of all mappings of a set A into the set $\mathbf{R}$ of all real numbers. For any $f, g \in X$, define $f \leq g$ if and only if $f(a) \leq g(a) \forall a \in A$. Show that $\leq$ is a partial ordering on X .

8. Let $f: A \rightarrow B$, $g: B \rightarrow C$ be two bijections. Prove that $(g \circ f)^{-1} = f^{-1} \circ g^{-1}$.

9. Given a set X , the set $\Delta = \{(x, x) \mid x \in X\}$ is called the *diagonal* of $X \times X$. For any relation R on X , the *inverse* of R is the relation

$$R^{-1} = \{(y, x) \mid (x, y) \in R\}$$

Prove the following:

(i) A relation R on X is reflexive if and only if $\Delta \subseteq R$.

(ii) A relation R on X is symmetric if and only if $R = R^{-1}$.

10. You know that a mapping $f: X \rightarrow Y$ is a relation. Give an example of a relation $f: X \rightarrow Y$ which is not mapping.

11. Let F be a set of sets. For any $A, B \in F$ define $A \leq B$ if and only if A is a subset of B . Show that $\leq$ is a partial ordering on F . (This partial ordering is called the *inclusion relation*)

12. Let X be the set of all lines. For any $a, b \in X$ define $a \sim b$ if and only if a is perpendicular to b . Show that $\sim$ is symmetric but neither reflexive nor transitive.

13. Let X be the set of all $n \times n$ matrices over the reals. For any $A \in X$, let $|A|$ denote its determinant. Given $A, B \in X$, define

Acc. no. - 16688

$A \sim B$ if and only if $|A| = |B|$. Verify that $\sim$ is an equivalence relation on X .

14. Let $X = \{1, 2, 3\}$.

(i) Find all relations on X .

(ii) Find all equivalence relations on X .

(iii) Find all partial orderings on X .

(iv) Find all those relations on X which are reflexive but neither symmetric nor transitive.

(v) Find all those relations on X which are symmetric but neither reflexive, nor transitive.

(vi) Find all those relations on X which are transitive, but neither reflexive nor symmetric.

(vii) Find all those relations on X which are neither reflexive, nor symmetric nor anti-symmetric, nor transitive.

15. Let $f: X \rightarrow Y$ be a given mapping. Prove the following:

(i) For any $B \subseteq Y$, $ff^{-1}(B) \subseteq B$. Further $ff^{-1}(B) = B \forall B \subseteq Y$ if and only if f is onto.

(ii) For any subset A of X , $A \subseteq f^{-1}f(A)$. Further $A = f^{-1}f(A) \forall A \subseteq X$ if and only if f is a one-to-one mapping.

(iii) $f(A_1 \cap A_2) = f(A_1) \cap f(A_2)$ holds for all subsets A_1, A_2 of X if and only if f is one-to-one.

(iv) For any subset A of X let A' denote its complement in X . Then $f(A') \supseteq f(A)'$, the complement of $f(A)$ in Y for every $A \subseteq X$ if and only if f is onto; $f(A') \subseteq f(A)'$ if and only if f is 1-1.

16. Prove the following:

(i) The relation of equipotence among sets in an equivalence relation.

(ii) If A, B, A', B' are any four sets such that $A \cap B = A' \cap B' = \phi$, $A \sim A', B \sim B'$ then $A \cup B \sim A' \cup B'$ and $A \times B \sim A' \times B'$.

(iii) For any two sets A, B let A^B denote the set of all mappings of B into A . Show that $A \sim A', B \sim B'$ implies $A^B \sim A'^{B'}$.

(iv) Show that a set A can never be equipotent to its power set.

17. Prove the following:

(i) Any subset of a countable set is countable.

(ii) Show that the mapping $f: \mathbf{Z} \rightarrow \mathbf{N}$ such that

$$f(n) = 2n \quad \forall n \geq 1, f(n) = -2n + 1 \quad \forall n \leq 0$$

is a bijection. Deduce that $\mathbf{Z}$ is denumerable.

(iii) Define $f: \mathbf{N} \times \mathbf{N} \rightarrow \mathbf{N}$ by $f(n, m) = 2^n \cdot 3^m$. Show that f is a one-to-one mapping and by using (i) show that $\mathbf{N} \times \mathbf{N}$ is denumerable.

(iv) Let $\{E_i\}_{i \in \mathbf{N}}$ be a countable family of countable sets. Show that its union is countable. [Hint: Use (iii)].

(v) Show that $\mathbf{Q}$ is denumerable.

(vi) Show that the set of all real numbers is not countable. Indeed show that $\mathbf{R} \sim P(\mathbf{N})$, the power set of $\mathbf{N}$.

Let $\{S_\alpha\}_{\alpha \in A}$ and $\{T_\beta\}_{\beta \in B}$ be any two families of sets. Prove the following:

$$(i) (\cup_{\alpha \in A} S_\alpha) \cap (\cap_{\beta \in B} T_\beta) = \cap_{\alpha, \beta} (S_\alpha \cap T_\beta).$$

(This is called the *generalized law of distributivity*.)

19. Let $f: X \rightarrow Y$, $g: Y \rightarrow Z$ and $h: Z \rightarrow X$ be three functions such that both hgf and gfh are 1-1 and fgh is onto. Prove that each of f , g and h is a 1-1 correspondence.

20. If a set S has n elements and let $\alpha: S \rightarrow S$ and $\beta: S \rightarrow S$ be such that $|\alpha(S)| = |\beta(S)| = n-1$, show that there exist permutations f and g on S such that $\beta = f \circ \alpha \circ g$.

21. Let k be an integer such that $1 \leq k \leq n$ and let X be a set having k elements and Y be a set having n elements. Prove that a number of injections from X to Y is equal to $\frac{n!}{(n-k)!}$.

22. Let m be an integer such that $1 \leq m \leq n$. Let $\sigma_n(m)$ be the number of surjections from a set having n elements to a set having m elements. Show that $\sigma_n(m) = m^n - \sum_{k=1}^{m-1} {}^m C_k \sigma_n(k)$.

Infer that

$$\sigma_n(m) = \sum_{k=0}^{m-1} (-1)^k {}^m C_{m-k} (m-k)^n = \sum_{j=1}^m (-1)^{m-j} {}^m C_j j^n$$

23. Let $g_{n,k}$ be the number of k -tuples $(a_1, \dots, a_k)$ of natural numbers such that $a_1 + a_2 + \dots + a_k = n$

(a) Show that $g_{n,1} = 1$ and that $g_{n,k} = \sum_{i=0}^n g_{i,k-1}$ for all $n \in \mathbf{N}$ and all $k \geq 2$.

(b) Show that $g_{n,k} = g_{n-1,k} + g_{n,k-1}$ for all $n \geq 1$ and for all $k \geq 2$.

(c) Conclude that

$$g_{n,k} = {}^{n+k-1} C_{k-1} \text{ for all } n \in \mathbf{N}, k \in \mathbf{N}.$$

24. A function $f: A \rightarrow B$ is said to be *left-cancellable* if for any set C and for any mappings g and h from C to A such that $fg = fh$ we have $g = h$.

Prove that a function $f: A \rightarrow B$ is left-cancellable if and only if f is 1-1.

[Hint: Let f be left-cancellable, and let $f(a) = f(b)$. Put $X = \{a\}$ and define $\lambda: X \rightarrow A$ by $\lambda(x) = a \forall x \in X$, and $\mu: X \rightarrow A$ by $\mu(x) = b \forall x \in X$. Show that $f\lambda = f\mu$.]

25. A function $f: A \rightarrow B$ is said to be *right-cancellable* if for any set C and for any mappings g and h from B to C such that $gf = hf$ we have $g = h$.

Prove that a function $f: A \rightarrow B$ is onto if and only if f is right-cancellable.

[Hint: Let f be right-cancellable.

Put $X = \{B - f(A)\} \cup \{\phi\}$ where ϕ is null set.

Define $\lambda: B \rightarrow X$ by $\lambda(b) = \phi$ when $b \in f(A)$

and $\lambda(b) = B - f(A)$ when $b \in B - f(A)$. Further define $\mu: B \rightarrow X$ by $\mu(b) = \phi \forall b \in B$. Show that $\lambda f = \mu f$.]

3. Binary Compositions

After having discussed relation and mapping, we now introduce the concept of a binary composition on a set. To start with consider the set $\mathbf{Z}$. There are three well-known operations in $\mathbf{Z}$ namely addition, multiplication and subtraction which, for every ordered pair (m, n) of elements of $\mathbf{Z}$ determine the elements $m+n$, mn , $m-n$ respectively of $\mathbf{Z}$. We can look upon addition, multiplication and subtraction as three mappings of $\mathbf{Z} \times \mathbf{Z}$ into $\mathbf{Z}$, which for each element (m, n) of $\mathbf{Z} \times \mathbf{Z}$ determine the elements $m+n$, mn , and $m-n$ respectively of $\mathbf{Z}$. One can construct many other mappings of $\mathbf{Z} \times \mathbf{Z}$ into $\mathbf{Z}$. All these mappings are examples of binary compositions which we define formally as under.

Definition 1.28. Given a set X , any mapping of $X \times X$ into X is called a *binary composition* on X .

Let f be a binary composition on X . By definition f is a mapping of $X \times X$ into X . Given $(a, b) \in X \times X$, $f(a, b) \in X$. For the sake of convenience, we shall write $a * b$ for $f(a, b)$, where $*$ is some appropriate fixed symbol and shall say that ' $*$ ' is a binary composition on X . Thus for every $(a, b) \in X \times X$, ' $*$ ' determines an element $a * b$ (unique, of course!) of X . The binary composition $*$ is called addition, multiplication or a circle composition according as we write '+', '·' or 'o' for $*$. If $a, b \in X$ and $a \neq b$ we know that $(a, b) \neq (b, a)$, so it is not necessary in general that the image in X of (a, b) and (b, a) under a binary composition $*$, must be equal. In other words it is not necessary that for a binary composition ' $*$ ' on X , $a * b = b * a$ must hold for all a, b in X . Some of the examples we shall discuss after a few definitions, will justify the remark made above.

Definition 1.29. Let ' $*$ ' be a binary composition on a set X . Then

(i) The binary composition ' $*$ ' is said to be *commutative* if $a * b = b * a$ for all $a, b \in X$.

(ii) The binary composition $*$ is said to be *associative* if $a * (b * c) = (a * b) * c$ for all $a, b, c \in X$.

(iii) An element e of X is said to be a *left identity* of X with respect to $*$ if $e * a = a$ for all $a \in X$.

(iv) An element f of X is said to be a *right identity* of X with respect to $*$ if $a * f = a$ for all $a \in X$.

(v) If an element i of X is a left as well as right identity then it is called an *identity* of X .

(vi) If e is an identity of X under $*$ and $a \in X$, then $a' \in X$ is called a *left or right inverse* of a according to $a' * a = e$ or $a * a' = e$. If an element a'' is both a right and left inverse of a , then a'' is called an *inverse* of a and a itself is said to be an *invertible element*.

Theorem 1.30. Let $*$ be a binary composition on a set X . Then the following hold:

(i) If X has a left identity and a right identity then they are equal.

(ii) If X has an identity (two-sided) then it is unique.

Proof: (i) Let e and f be left and right identity respectively, of X . Then $e * f = f$, since e is left identity.
and $e * f = e$, since f is a right identity.
Consequently $e = f$.

For second part, let e_1 and e_2 be two identities of X . Then since e_1 is a left identity and e_2 , a right identity; by first part we have $e_1 = e_2$. ■

EXAMPLE 20. For any three integers m, n, p we know that

$$(i) \quad m + n = n + m$$

$$(ii) \quad m + (n + p) = (m + n) + p$$

$$(iii) \quad m + 0 = m = 0 + m$$

(iv) $m + (-m) = 0 = (-m) + m$ and $-m$ is an integer. Here (i) and (ii) show that addition is commutative and associative binary composition on $\mathbf{Z}$. (iii) shows that 0 is the identity of $\mathbf{Z}$ under addition. (iv) show that for each $m \in \mathbf{Z}$, $-m$ is its inverse under addition.

EXAMPLE 21. For any three integers m, n, p we know that

$$(i) \quad m \cdot n = n \cdot m$$

$$(ii) \quad m \cdot (n \cdot p) = (m \cdot n) \cdot p$$

$$(iii) \quad m \cdot 1 = m = 1 \cdot m$$

This shows that the usual multiplication composition on $\mathbf{Z}$ is commutative, associative and 1 is the identity of $\mathbf{Z}$ with respect to multiplication. Now $(-1) \cdot (-1) = 1$ shows that -1 is its own inverse under multiplication. We know that for any integer m , different from 1 and -1 , there exists no integer n such that $m \cdot n = 1$. Thus m has no inverse in $\mathbf{Z}$ under multiplication.

EXAMPLE 22. We know that for any two integers m and n , $m \cdot (-n)$

is again an integer. Thus the operation of subtraction is a binary composition on $\mathbf{Z}$. Now $2, 3, 5 \in \mathbf{Z}$ and

$$(i) \quad 2-3 \neq 3-2$$

$$(ii) \quad 2-(3-5) \neq (2-3)-5$$

$$(iii) \quad 2-0 = 2 \neq 0-2$$

(i) and (ii) show that subtraction is neither commutative nor associative. Since for any integer k , $k-0=k$, 0 is right identity of $\mathbf{Z}$ under $'-'$, but (iii) shows 0 is not a left identity under $'-'$. So that $\mathbf{Z}$ has right identity but no left identity under subtraction.

EXAMPLE 23. Let X be any set having at least three elements and let $S(X)$ denote the set of all permutations of X . Theorem 1.17 shows that the resultant composition $'o'$ is associative, the identity map i on X is the identity and each element of $S(X)$ is invertible. Let a, b, c be three distinct elements of X . Define $f: X \rightarrow X$ and $g: X \rightarrow X$ as follows:

$$f(a)=b, f(b)=a \text{ and } f(x)=x \quad \forall x \in X \text{ different from } a \text{ and } b.$$

$$g(a)=b, g(b)=c, g(c)=a \text{ and } g(x)=x \quad \forall x \neq a, b, c.$$

Then $f, g \in S(X)$.

$$\text{Now } (gof)(a)=g[f(a)]=g(b)=c$$

$$\text{and } (fog)(a)=f(g(a))=f(b)=a.$$

$$\text{Thus } (fog)(a) \neq (gof)(a).$$

Consequently $fog \neq gof$.

Hence the resultant composition on $S(X)$ is not commutative.

EXAMPLE 24. Let M_2 be the set of all 2×2 matrices whose elements are rational numbers. For any two matrices

$$A = \begin{bmatrix} a_1 & b_1 \\ c_1 & d_1 \end{bmatrix}, \quad B = \begin{bmatrix} a_2 & b_2 \\ c_2 & d_2 \end{bmatrix}$$

the usual sum and product

$$A+B = \begin{bmatrix} a_1+a_2 & b_1+b_2 \\ c_1+c_2 & d_1+d_2 \end{bmatrix} \quad AB = \begin{bmatrix} a_1a_2+b_1c_2 & a_1b_2+b_1d_2 \\ c_1a_2+d_1c_2 & c_1b_2+d_1d_2 \end{bmatrix}$$

are again in M_2 . So that usual matrix addition and multiplication are two binary compositions on M_2 . From the well-known properties of matrices we know that addition in M_2 is commutative and associative. The matrix

$$\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

is the identity under addition and

$$-A = \begin{bmatrix} -a_1 & -b_1 \\ -c_1 & -d_1 \end{bmatrix}$$

is the inverse of A under addition. Same way multiplication is associative.

$$I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

is the identity under multiplication. Only non-singular matrices have inverse under multiplication.

$$\text{Since for } A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \quad B = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$$

$AB \neq BA$, multiplication is not even commutative.

Composition tables for finite sets

Let X be a finite non-void set and $*$ be a binary composition on X . The composition table for $*$ is to be constructed as follows:

Let $X = \{x_1, x_2, \dots, x_n\}$.

We write the elements $x_1, x_2, \dots, x_n$ of the set of X in a horizontal row as well as in a vertical column. Then put down the element $x_i * x_j$ at the intersection of the row headed by x_i ($1 \leq i \leq n$) and the column headed by x_j ($1 \leq j \leq n$). Then we get the following table.

	x_1	x_2	x_3	...	x_i	...	x_j	...	x_n
x_1	$x_1 * x_1$	$x_1 * x_2$	$x_1 * x_3$	...	$x_1 * x_i$	...	$x_1 * x_j$	...	$x_1 * x_n$
x_2	$x_2 * x_1$	$x_2 * x_2$	$x_2 * x_3$	...	$x_2 * x_i$	...	$x_2 * x_j$	...	$x_2 * x_n$
x_3	$x_3 * x_1$	$x_3 * x_2$	$x_3 * x_3$	...	$x_3 * x_i$	...	$x_3 * x_j$	...	$x_3 * x_n$
...	...	...	...	...	...	...	...	...	...
x_i	$x_i * x_1$	$x_i * x_2$	$x_i * x_3$	...	$x_i * x_i$	...	$x_i * x_j$	...	$x_i * x_n$
...	...	...	...	...	...	...	...	...	...
x_j	$x_j * x_1$	$x_j * x_2$	$x_j * x_3$	...	$x_j * x_i$	...	$x_j * x_j$	...	$x_j * x_n$
...	...	...	...	...	...	...	...	...	...
x_n	$x_n * x_1$	$x_n * x_2$	$x_n * x_3$	...	$x_n * x_i$	...	$x_n * x_j$	...	$x_n * x_n$

EXAMPLE 25. Let $X = \{0, 1, 2, 3\}$. Define a composition $*$ on X as follows: Given $a, b \in X$ define $a * b = c$, where c is the least non-negative remainder got when $a + b$ is divided by 4. For example

$1 * 3 = 0, 2 * 3 = 1, 3 * 3 = 2$. Following is the composition table for the binary composition $*$ on X

	0	1	2	3
0	0	1	2	3
1	1	2	3	0
2	2	3	0	1
3	3	0	1	2

EXAMPLE 26. Let $X = \{0, 1, 2, 3\}$. Given $a, b \in X$ define $a * b = a$. For example $2 * 3 = 2, 3 * 2 = 3$. Following is the composition table for the composition $*$

	0	1	2	3
0	0	0	0	0
1	1	1	1	1
2	2	2	2	2
3	3	3	3	3

We now come back to our old friend, the set of integers $\mathbf{Z}$. Let E be the set of all even integers. Then $E < \mathbf{Z}$. We know that if m and n are even integers, then $m+n$ is also an even integer. In other words given $m, n \in E$, then $m+n \in E$. We say that E is a closed subset of $\mathbf{Z}$ with respect to (or under) addition composition on $\mathbf{Z}$. We define closed subset of a set under a binary composition.

Definition 1.31. Let X be a non-void set and $*$ be a binary composition on X . A subset H of X is said to be *closed* under the binary composition $*$, if $a * b \in H$ for all $a, b \in H$.

Let X be a non-void set with a binary composition $*$ and A , a subset of X closed under the binary composition $*$. Since for any $a, b \in A$, $a * b \in A$, we get a mapping $f: A \times A \rightarrow A$ such that $f(a, b) = a * b$ for all $a, b \in A$. Thus by definition, f is a binary composition on A . This binary composition f on A is said to be *induced* by $*$, or simply an *induced binary composition*.

EXAMPLE 27. The set E of all even integers is a closed subset of $\mathbf{Z}$ under addition composition.

EXAMPLE 28. Let K be the set of all odd integers. K is a subset of $\mathbf{Z}$. However the sum of two odd integers is an even integer. Thus if $m \in K, n \in K$, then $m+n \notin K$. Consequently K is not closed under addition. However K is a closed subset of $\mathbf{Z}$ under multiplication.

Example 28 shows that if set X has more than one binary composition, then a subset A of X may be closed under one binary composition but may not be closed under the other binary composition.

PROBLEMS

1. If a set A has n members, then what is the number of binary compositions on A ? (Ans. n^{n^2})

2. For any $a, b \in \mathbf{Q}$, define $a * b = \frac{ab}{2}$. Show that $\mathbf{Z}$ is not closed under $*$, however E , set of all even integers is closed under $*$. Verify further that $*$ is associative, commutative, and 2 is the identity for $*$. Further every non-zero element of $\mathbf{Q}$ is invertible under $*$.

3. For any two integers m, n define $m * n = m + n + 1$ show the following:

(i) E , the set of all even integers is not closed under $*$

(ii) K , the set of all odd integers is closed under $*$.

(iii) $*$ is associative, commutative. Identity with respect to $*$ exists in $\mathbf{Z}$ and every element of $\mathbf{Z}$ is invertible under $*$.

4. Let M_2 be the set of all 2×2 matrices over the rational numbers. For any $A, B \in M_2$, define

$$A \circ B = \frac{AB + BA}{2}.$$

Prove the following

(i) $\circ$ is commutative

(ii) For $A = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$, $B = \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix}$,

$$C = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}$$

$(A \circ B) \circ C \neq A \circ (B \circ C)$. This shows that $\circ$ is not associative.

(iii) $I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

is the identity under ' $\circ$ '.

5. For any two matrices A, B in M_2 , define

$$A * B = \frac{AB - BA}{2}.$$

Prove the following:

(i) For any $A \in M_2$, $A * A = O$, where O is zero matrix.

(ii) For $A, B, C \in M_2$,

$$A * (B * C) + B * (C * A) + C * (A * B) = O.$$

(iii) There exists no identity in M_2 with respect to $*$.

(iv) For any $A, B, C \in M_2$,

$$A * (B + C) = A * B + A * C.$$

$$(B + C) * A = B * A + C * A.$$

(v) Show that $*$ is neither commutative nor associative.

6. Let $\mathbf{Z}[i]$ denote the set of all complex numbers of the form $a + bi$, where a, b are integers. Show the following:

(i) $\mathbf{Z}[i]$ is closed under usual addition and multiplication of complex numbers. $\mathbf{Z}[i]$ contains an identity with respect to addition and an identity with respect to multiplication. Show that these two identities are distinct.

(ii) Find all those elements of $\mathbf{Z}[i]$ which are invertible under multiplication.

7. For any two sets A and B , the set $(A - B) \cup (B - A)$ is called the *symmetric differences* of A and B and is denoted by $A \nabla B$. Let X be any set and $P(X)$ be its power set, prove the following:

(i) For any $A, B \in P(X)$, $A \nabla B \in P(X)$.

For any $A, B, C \in P(X)$,

(ii) $A \nabla B = B \nabla A$.

(iii) $(A \nabla B) \nabla C = B \nabla (B \nabla C)$.

(iv) $A \nabla B = A$ if and only if $B = \phi$ (**Poretsky's Law**).

(v) $A \nabla A = \phi$.

(vi) $A \cap (B \nabla C) = (A \cap B) \nabla (A \cap C)$.

[Hint for (iii) : $(A \nabla B) \nabla C = [A - (B \cup C)] \cup [B - (C \cup A)] \cup [C - (A \cup B)] \cup [A \cap B \cap C]$.

8. For any sets A_1, A_2, B_1 and B_2 . Prove that

$$(A_1 \cup A_2) \nabla (B_1 \cup B_2) \subseteq (A_1 \nabla B_1) \cup (A_2 \nabla B_2)$$

9. Show that there exists no binary system with identity and two-sided inverses in which the associative law fails for just one ordered triple.

10. A set A has n elements, what is the number of commutative binary composition on A ?

$$\left(\text{Ans. } n \frac{n^2 + n}{2} \right)$$

4. Some Elementary Results of Number Theory

In the chapters to follow, a frequent use is made of certain results of number theory. Since most of these results are not taught in an elementary course on algebra, we now give a brief resume' of that portion of Number theory which will be needed by us in subsequent part of the book.

REMARK. Throughout this section, all symbols written in lower-case Latin letters will stand for integers.

Definition 1.32. A non-zero integer b is said to divide a if $a = bk$ for some k .

NOTATION. If b divides a , we write $b \mid a$.

REMARK. Following statements are equivalent:

' b divides a ', ' b is a divisor of a ', ' b is a factor of a ', ' a is a multiple of b '.

Theorem 1.33. Let a be a non-zero integer, then

- (i) $a \mid 0$.
- (ii) $a \mid a$.
- (iii) $a \mid b$, then for each c , $a \mid bc$.
- (iv) $a \mid b$ and $a \mid c$, then $a \mid (bx + cy)$ for all x, y .
- (v) $a \mid b$ and $b \mid c$ ($b \neq 0$), then $a \mid c$.
- (vi) $a \mid b$ and $b \mid a$ ($b \neq 0$), then $a = \pm b$.

Proof: (i) $0 = a \cdot 0$.

(ii) $a = a \cdot 1$.

(iii) $a \mid b \Rightarrow b = ak$ for some $k \Rightarrow bc = a(kc) \Rightarrow a \mid bc$.

(iv) $a \mid b \Rightarrow b = ak$ for some k ; $a \mid c \Rightarrow c = al$ for some l .

Thus $a \mid b$ and $a \mid c \Rightarrow b = ak$ and $c = al$.

This in turn gives

$$bk + cy = akx + aly = a(kx + ly) \Rightarrow a \mid (bx + cy).$$

(v) $a \mid b \Rightarrow b = ak$ for some k

$$b \mid c \Rightarrow c = bl \text{ for some } l \Rightarrow c = (ak)l = a(kl) \Rightarrow a \mid c.$$

(vi) $a \mid b \Rightarrow b = ak$ for some k

$$b \mid a \Rightarrow a = bl \text{ for some } l$$

$$\text{Thus } b = ak = blk \Rightarrow lk = 1$$

$$\text{Hence } l = \pm 1, \text{ i.e. } a = \pm b. \blacksquare$$

Before starting next theorem, we recall the principle of well-ordering.

Principle of well-ordering. Every non-empty set of real numbers which is bounded below must possess a least element.

As a particular case we get "If X is a subset of integers and y is a positive integer in X , then X contains least positive integer."

Following theorem is one of the fundamental results of Number theory. This is commonly known as *Division Algorithm* or sometimes

as Euclidean Algorithm.

Theorem 1.34. If a and b are integers, $b \neq 0$, then there exist integers q and r , such that $a = bq + r$ with $0 \leq r < |b|$. Further if $a = bq_1 + r_1$ with $0 \leq r_1 < |b|$; then $r = r_1$ and $q = q_1$.

Proof: Firstly we suppose that $b > 0$, then $|b| = b$. Three cases arise:

Case I. $a \geq b$. If $a = bk$ for some k we have $q = k$ and $r = 0$. So the theorem holds in this case.

If, there exists no integer k such that $a = bk$, consider the set $X = \{a - bt \mid t \in \mathbb{Z}\}$. X is non-empty as $a - b \cdot 0 = a \in X$ and since $a > 0$, X has at least one positive integer. By principle of well-ordering, X has a least positive integer as its member. Let it be $r = a - bq$. Now $r = b$ is not possible otherwise we shall get $a = b(q+1)$ which is against the assumption of this case. Suppose then $r > b$. This implies that $r - b > 0$ i.e., $a - b(q+1) > 0$. Clearly $a - b(q+1) \in X$. Further $(a - bq) - (a - b(q+1)) = b > 0$ gives $a - b(q+1) < a - bq$, which is against the choice of r . So $r < b$.

Hence $a = bq + r$ with $0 \leq r < b$.

Case II. $0 \leq a < b$.

We write $a = b \cdot 0 + a$.

Case III. $a < 0$. This gives $-a > 0$ and so by case I and case II, $-a = bq' + s$ with $0 \leq s < b$. Thus we get $a = b(-q') + (-s)$. If $s = 0$, $-s = 0$ and the result holds trivially. If $s > 0$, then we can write $a = b(-q' - 1) + (b - s)$ with $0 < b - s < b$. Hence $a = bq + r$ with $q = -q' - 1$, $r = b - s$.

Secondly we discuss the case $b < 0$. Here $|b| = -b$. Since $-b > 0$, by our earlier discussion we can find q_1 and r such that $a = (-b)q_1 + r$ with $0 \leq r < -b$. This can be written as $a = b(-q_1) + r$. Taking $q = -q_1$, we get $a = bq + r$ with $0 \leq r < |b|$.

Thus the first statement of the theorem is proved. For uniqueness part, let $r > r_1$. This means $q \neq q_1$.

Then $a = bq + r = bq_1 + r_1 \Rightarrow b(q - q_1) = r_1 - r$.

$\Rightarrow |b| \mid |q - q_1| = |r_1 - r| = r - r_1$. As $|q - q_1| \geq 1$,

$r - r_1 \geq |b| \Rightarrow r \geq r_1 + |b| \Rightarrow r \geq |b|$

which is against first part.

So $r \leq r_1$. Similarly it can be proved that $r_1 \leq r$.

Hence $r = r_1$ and as a consequence of this $bq = bq_1 \Rightarrow q = q_1$.

This completes the proof of the theorem. ■

Definition 1.35. A positive integer c is said to be *Highest Common Factor (HCF)* of two non-zero integers a and b if

(i) $c \mid a$ and $c \mid b$

(ii) If $d \mid a$ and $d \mid b$, then $d \mid c$.

Definition 1.36. A positive integer g is called *Least Common Multiple* (LCM) of two non-zero integers a and b , if

- (i) $a \mid g$ and $b \mid g$.
- (ii) If $a \mid h$ and $b \mid h$, then $g \mid h$.

NOTATION. HCF of a and b is denoted by (a, b) .

LCM of a and b is denoted by $[a, b]$.

REMARK. If HCF of two integers a and b exists then it must be unique. For suppose c' is another HCF of a and b . Then by (ii) of definition 1.35

$$c \mid c' \text{ and } c' \mid c \Rightarrow c = \pm c' \quad (\text{Theorem 1.33})$$

$$\Rightarrow c = c' \text{ as both } c \text{ and } c' \text{ are positive.}$$

Similarly it can be shown that if LCM of two numbers a and b exists, it must be unique.

EXAMPLE 29. 6 is HCF of 12 and 18.

All the factors of 12 are $\pm 1, \pm 2, \pm 3, \pm 4, \pm 6, \pm 12$, while all the factors of 18 are $\pm 1, \pm 2, \pm 3, \pm 6, \pm 9, \pm 18$.

Then clearly $6 \mid 12$ and $6 \mid 18$ and every common factor of 12 and 18 is also a factor of 6.

EXAMPLE 30. 36 is LCM of 12 and 18.

Clearly $12 \mid 36$ and $18 \mid 36$.

If $12 \mid h$ and $18 \mid h$, then $4 \mid h$ and $9 \mid h$. Then $4 \cdot 9 = 36 \mid h$.

[Cf. Corollary 2, Theorem 1.39, as $(4, 9) = 1$].

Following theorem shows the existence of HCF for every pair of non-zero integers.

Theorem 1.37. Each pair of non-zero integers a and b has HCF.

Proof: Let $X = \{ta + ub \mid t, u \in \mathbb{Z}\}$.

Clearly $a = 1a + 0b \in X$. So X contains at least a non-zero integer. Further if $\alpha = ta + ub \in X$ is negative integer then $-\alpha = (-t)a + (-u)b$ is positive integer in X .

Thus X is non-empty subset of integers having at least one positive integer.

By principle of well-ordering there exists $c = xa + yb$, the least positive integer belonging to X . We claim $c = (a, b)$.

By division algorithm for any $ka + lb \in X$, $ka + lb = cq + r$ with $0 \leq r < c$. If $r \neq 0$, then

$$r = ka + lb - cq = ka + lb - (xaq + ybq) = (k - xq)a + (l - yq)b$$

is a positive integer belonging to X .

As $r < c$, we get a contradiction to the choice of c . Hence $r = 0$.

In other words $c \mid ka + lb$. Thus c is a factor of every element of set X . As $a = 1a + 0b$ and $b = 0a + 1b$ are elements of X , we get $c \mid a$ and $c \mid b$.

Suppose now $d \mid a$ and $d \mid b$, then $d \mid (xa + yb)$, (Theorem 1.33).

i.e., $d \mid c$. Hence $c = (a, b)$. ■

Definition 1.38. Two non-negative integers a and b are said to be co-prime or relatively prime if $(a, b) = 1$.

EXAMPLE 31. 4 and 9 are co-prime as factors of 4 are $\pm 1, \pm 2, \pm 4$ while those of 9 are $\pm 1, \pm 3, \pm 9$.

Hence HCF of 4 and 9 is 1.

Theorem 1.39. Two non-zero integers a and b are co-prime if and only if there exist x, y such that $ax + by = 1$. (Bezout's Identity).

Proof: Let $(a, b) = 1$, then by Theorem 1.37, $1 = ax + by$ for some x and y .

Conversely let $ax + by = 1$ for some x, y .

Suppose $c = (a, b)$. Then $c \mid a$ and $c \mid b$. By Theorem 1.33, $c \mid (ax + by) \Rightarrow c \mid 1$.

As $c > 0$, this immediately gives $c = 1$. Hence $(a, b) = 1$. ■

Corollary 1. If $a \mid bc$ and $(a, b) = 1$, then $a \mid c$.

Proof: As $(a, b) = 1 \Rightarrow ax + by = 1$ for some x, y we get
 $acx + bcy = c$.

Now $a \mid a, a \mid bc$, so by Theorem 1.33, $a \mid [a(cx) + (bc)y]$ i.e. $a \mid c$. ■

Corollary 2. If $a \mid b, c \mid b$ and $(a, c) = 1$, then $ac \mid b$.

Proof: As $(a, c) = 1, ax + cy = 1$ for some x, y .

Now $a \mid b \Rightarrow b = ak$ for some k . Thus we get $akx + kcy = k$ i.e. $bx + kcy = k$.

As $c \mid b$ and $c \mid c, c \mid (xb + kcy)$, (Theorem 1.33).

This implies $c \mid k$ or $k = cd$ for some d .

Hence $b = ak = acd \Rightarrow ac \mid b$. ■

Definition 1.40. An integer $p > 1$ is called a prime (prime number) if the only factors of p are $\pm 1, \pm p$.

EXAMPLE 32. 2, 3, 5, 7, 11, 13, 17, 19 are all primes, while 4, 6, 8, 10 are not prime.

Definition 1.41. An integer $n > 1$ is called a composite if it is not a prime.

In other words an integer $n > 1$ is called composite if n can be factored as ab with $a > 1$ and $b > 1$.

Theorem 1.42. If p is a prime number then for any integer n , either $(n, p) = 1$ or $(n, p) = p$.

Proof: Suppose $(n, p) = d$ then $d \mid p$ and $d \mid n$.

Now d is positive and the only positive factors of p are 1 and p . Hence the theorem follows. ■

Corollary. If p is a prime number such that $p \mid ab$, then either $p \mid a$ or $p \mid b$.

Proof: By above theorem if $p \nmid a$ then $(p, a) = 1$.

Invoking Corollary 1 of Theorem 1.39, we get $p \mid b$. ■

We are now all set for the so-called 'Fundamental Theorem of Arithmetic'.

Theorem 1.43. (Fundamental Theorem of Arithmetic).

If n is a positive integer greater than 1, then n can be written uniquely as $p_1^{k_1} p_2^{k_2} \dots p_t^{k_t}$, where $p_1 > p_2 > \dots > p_t$ are prime numbers and for all i , $1 \leq i \leq t$, k_i are integers > 0 .

Proof: We shall apply induction on n .

For $n=2$, the theorem trivially holds as we can take $t=1$, $k_1=1$, $p_1=2$.

Let the theorem be true for all integers m , $2 \leq m < n$.

Now if n is a prime number then again the theorem holds as we can take $t=1$, $k_1=1$ and $p_1=n$. If n is not a prime number then we can write $n=ab$ with $a > 1$, $b > 1$.

As $a > 1$, $b < n$; similarly $a < n$. By induction hypothesis $a=p_1^{\alpha_1} \dots p_r^{\alpha_r}$ and $b=q_1^{\beta_1} \dots q_s^{\beta_s}$ with p_i and q_j are prime numbers and α_i, β_j are integers > 0 for all $1 \leq i \leq r$, $1 \leq j \leq s$. Hence the theorem holds for n .

For the uniqueness we apply a similar technique.

Suppose that $n=p_1^{k_1} p_2^{k_2} \dots p_t^{k_t}=q_1^{s_1} q_2^{s_2} \dots q_r^{s_r}$ where $p_1 > p_2 > \dots > p_t$; $q_1 > q_2 > \dots > q_r$ and k_i, s_j are integers > 0 for all $1 \leq i \leq t$, $1 \leq j \leq r$.

The result is true for $n=2$. Suppose this is true for all integers m , $2 \leq m < n$.

Now $p_1^{k_1} p_2^{k_2} \dots p_t^{k_t}=q_1^{s_1} q_2^{s_2} \dots q_r^{s_r}$ and $k_1 > 0$, so p_1 divides left hand side. This implies $p_1 \mid q_1^{s_1} q_2^{s_2} \dots q_r^{s_r}$

By Corollary to Theorem 1.42, there exists an i , $1 \leq i \leq r$ such that $p \mid q_i$. As q_i is prime this gives $p_1=q_i$.

Now $q_1 \geq q_i \Rightarrow q_1 \geq p_1$. In the same way we get $q_1=p_j$ for some $1 \leq j \leq t$ and $p_1 \geq q_1$. As a consequence $p_1=q_1$.

If $k_1 > s_1$, then after cancelling p_1 from sides we get

$$p_1^{k_1-s_1} p_2^{k_2} \dots p_t^{k_t} = q_2^{s_2} \dots q_r^{s_r}.$$

This will in turn imply, by the arguments given earlier, that $p_1=q_2$. This is absurd as $q_1 > q_2$. Hence $k_1 \leq s_1$. Similarly it can be proved that $s_1 \leq k_1$. Hence $k_1=s_1$. So we can cancel $p_1^{k_1}=q_1^{s_1}$ from both sides to obtain $p_2^{k_2} \dots p_t^{k_t}=q_2^{s_2} \dots q_r^{s_r}$.

If $c=p_2^{k_2} \dots p_t^{k_t}=q_2^{s_2} \dots q_r^{s_r}=1$ then $k_2=\dots=k_t=s_2=\dots=s_r=0$, this is absurd as each $k_i > 0$, $s_j > 0$.

Hence $c > 1$. Evidently $c < n$. So by induction hypothesis $r-1=t-1$ and $k_i=s_r$, $p_i=q_i \forall 2 \leq i \leq r-1$. Thus $r=t$ and $p_j=q_j$, $k_j=s_j$ for all $1 \leq j \leq r$. Hence the theorem is completely proved. ■

Corollary. The set of prime numbers is not finite.

Proof: Suppose that the set of prime numbers is finite. So it can be put into 1-1 correspondence with the set $\{1, 2, 3, \dots, n\}$ for some natural number n .

Put $f(i) = p_i$ for all $1 \leq i \leq n$. Then $\{p_1, p_2, \dots, p_n\}$ is the set of all prime numbers.

Consider the integer $p_1 p_2 p_3 \dots p_n + 1$.

As $p_1 p_2 p_3 \dots p_n + 1 > 1$, by above theorem there exists a prime number q such that $q \mid (p_1 p_2 p_3 \dots p_n + 1)$.

Now all primes must be in the set $\{p_1, p_2, \dots, p_n\}$, so $q = p_i$ for some $1 \leq i \leq n$. This implies $q \mid p_1 p_2 \dots p_n$. As a consequence $q \mid 1$, which is absurd. Hence our contention is established.

Definition 1.44. For any integer $n \geq 1$, the number of all integers i , $1 \leq i \leq n$ such that $(n, i) = 1$ is called *Euler's ϕ -function* (or *Totient function*) and is denoted by $\phi(n)$.

EXAMPLE 33. $\phi(1) = 1$, $\phi(2) = 1$, $\phi(3) = 2$, $\phi(4) = 2$.

To compute $\phi(12)$, we consider 1, 2, 3, 4, ..., 12.

Starting with 1 we note that $(1, 12) = 1$; $(2, 12) = 2$, $(3, 12) = 3$; $(4, 12) = 4$; $(5, 12) = 1$; $(6, 12) = 6$; $(7, 12) = 1$; $(8, 12) = 4$; $(9, 12) = 3$; $(10, 12) = 2$; $(11, 12) = 1$ and $(12, 12) = 12$.

Thus only 1, 5, 7, 11 fulfil the condition of our definition. Hence $\phi(12) = 4$.

In the following we give a formula which makes the computation of $\phi(n)$ very easy for every $n > 1$.

Theorem 1.45. If $p_1, p_2, \dots, p_k$ are distinct prime factors of n then

$$\phi(n) = n \left(1 - \frac{1}{p_1} \right) \left(1 - \frac{1}{p_2} \right) \dots \left(1 - \frac{1}{p_k} \right).$$

Proof: We can write $n = p_1^{\alpha_1} p_2^{\alpha_2} \dots p_k^{\alpha_k}$ with $\alpha_i > 0$ for all $1 \leq i \leq k$.

Let $1 < m \leq n$. Then if $(m, n) = d$, $d > 1$ if and only if one of p_i 's, $1 \leq i \leq k$, is a factor of d . This gives that $\phi(n)$ = number of integers lying between 1 and n which do not have any of $p_1, p_2, \dots, p_k$ as a factor.

Now we apply induction on k .

For $k = 1$, let $X = \{1, 2, \dots, n\}$.

Out of the integers in X , only $p_1, 2p_1, 3p_1, \dots, \left(\frac{n}{p_1}\right)p_1$ are divisible by p_1 . These are $\frac{n}{p_1}$ in number. Therefore

$$\phi(n) = n - \frac{n}{p_1} = n \left(1 - \frac{1}{p_1} \right).$$

Hence theorem holds for $k=1$.

Let now $k > 1$ and theorem holds for number of distinct prime factors of n , less than k .

So the number of integers lying between 1 and n , which do not have any of $p_1, p_2, \dots, p_{k-1}$ as a factor is given by

$$n \left(1 - \frac{1}{p_1} \right) \left(1 - \frac{1}{p_2} \right) \dots \left(1 - \frac{1}{p_{k-1}} \right).$$

Again, out of the integers in X , only $p_k, 2p_k, \dots, \left(\frac{n}{p_k} \right) p_k$ are divisible by p_k . Put $Y = \left\{ p_k, 2p_k, \dots, \left(\frac{n}{p_k} \right) p_k \right\}$. Among the integers in Y , the number of integers not having the factors $p_1, p_2, \dots, p_{k-1}$ is equal to the number of coefficients $1, 2, \dots, \left(\frac{n}{p_k} \right)$, co-prime to $p_1, \dots, p_{k-1}$; as $(p_k, p_j) = 1$ for all $j \neq k$. By induction hypothesis this number is equal to

$$\frac{n}{p_k} \left(1 - \frac{1}{p_1} \right) \left(1 - \frac{1}{p_2} \right) \dots \left(1 - \frac{1}{p_{k-1}} \right).$$

Also by induction hypothesis, the number of integers in X not having $p_1, p_2, \dots, p_k$ is equal to

$$n \left(1 - \frac{1}{p_1} \right) \left(1 - \frac{1}{p_2} \right) \dots \left(1 - \frac{1}{p_{k-1}} \right).$$

Among these, those integers are also taken into account which are divisible by p_k . But the number of such integers is

$$\frac{n}{p_k} \left(1 - \frac{1}{p_1} \right) \left(1 - \frac{1}{p_2} \right) \dots \left(1 - \frac{1}{p_{k-1}} \right).$$

Thus the total number of integers lying between 1 and n and not having any of the factors $p_1, p_2, \dots, p_k$ is given by

$$n \left(1 - \frac{1}{p_1} \right) \left(1 - \frac{1}{p_2} \right) \dots \left(1 - \frac{1}{p_{k-1}} \right) - \frac{n}{p_k} \left(1 - \frac{1}{p_1} \right) \left(1 - \frac{1}{p_2} \right) \dots \left(1 - \frac{1}{p_{k-1}} \right) = n \left(1 - \frac{1}{p_1} \right) \left(1 - \frac{1}{p_2} \right) \dots \left(1 - \frac{1}{p_k} \right). \blacksquare$$

Corollary. If a and b are integers ≥ 1 such that $(a, b) = 1$, then $\phi(ab) = \phi(a) \phi(b)$.

Proof: If any of a, b is 1, the result is obviously true as $\phi(1) = 1$.

So let $a > 1$ and $b > 1$. Then $a = p_1^{\alpha_1} \dots p_k^{\alpha_k}$ and $b = q_1^{\beta_1} \dots q_s^{\beta_s}$ where all $1 \leq i \leq k$, $1 \leq j \leq s$, p_i, q_j are prime numbers and $\alpha_i > 0, \beta_j > 0$. As $(a, b) = 1$, $p_1, p_2, \dots, p_k, q_1, q_2, \dots, q_s$ are all distinct prime numbers.

Now $ab = p_1^{\alpha_1} \dots p_k^{\alpha_k} q_1^{\beta_1} \dots q_s^{\beta_s}$.

By above theorem

$$\begin{aligned}
\phi(ab) &= p_1^{\alpha_1} \dots p_k^{\alpha_k} q_1^{\beta_1} \dots q_s^{\beta_s} \left(1 - \frac{1}{p_1}\right) \left(1 - \frac{1}{p_2}\right) \dots \left(1 - \frac{1}{p_k}\right) \\
&\quad \left(1 - \frac{1}{q_1}\right) \left(1 - \frac{1}{q_2}\right) \dots \left(1 - \frac{1}{q_s}\right) \\
&= p_1^{\alpha_1} p_2^{\alpha_2} \dots p_k^{\alpha_k} \left(1 - \frac{1}{p_1}\right) \left(1 - \frac{1}{p_2}\right) \dots \left(1 - \frac{1}{p_k}\right) \\
&\quad \times q_1^{\beta_1} q_2^{\beta_2} \dots q_s^{\beta_s} \left(1 - \frac{1}{q_1}\right) \left(1 - \frac{1}{q_2}\right) \dots \left(1 - \frac{1}{q_s}\right) \\
&= \phi(a) \phi(b).
\end{aligned}$$

EXAMPLE 34. $36 = 2^2 \cdot 3^2$ so $\phi(36) = 36(1 - \frac{1}{2})(1 - \frac{1}{3}) = 12$.

Directly one can check that only 1, 5, 7, 11, 13, 17, 19, 23, 25, 29, 31, 35 are co-prime to 36 and are less than 36.

REMARK. The corollary given above is not true if $(a, b) \neq 1$. For example take $a=4, b=6$.

Then $ab=24$. $\phi(a)=\phi(4)=2$, $\phi(b)=\phi(6)=2$ and $\phi(ab)=\phi(24)=8$ but $8 \neq 2 \cdot 2$.

Theorem 1.46. If $n \geq 1$ and $d \geq 1$ then $\sum_{d|n} \phi(d) = n$

Proof: If $n=1$, then the only positive factor of 1 is 1 so we get $\sum_{d|1} \phi(d) = \phi(1) = 1 = n$.

If $n > 1$; let us first assume that $n = p_1^{\alpha_1}$, $\alpha_1 > 0$

The only positive factors of n are 1, p_1 , $p_1^2, \dots, p_1^{\alpha_1}$

$$\begin{aligned}
\text{so } \sum_{d|n} \phi(d) &= \phi(1) + \phi(p_1) + \phi(p_1^2) + \dots + \phi(p_1^{\alpha_1}) \\
&= 1 + p_1 \left(1 - \frac{1}{p_1}\right) + p_1^2 \left(1 - \frac{1}{p_1}\right) + \dots + p_1^{\alpha_1} \left(1 - \frac{1}{p_1}\right) \\
&= 1 + (p_1 - 1) + (p_1^2 - p_1) + \dots + (p_1^{\alpha_1} - p_1^{\alpha_1-1}) \\
&= p_1^{\alpha_1} = n.
\end{aligned}$$

Let $n = p_1^{\alpha_1} p_2^{\alpha_2} \dots p_k^{\alpha_k}$ with $\alpha_k > 0$ and let the result be true for all integers ≥ 1 having $k-1$ distinct prime numbers as factors.

We can write $n = q p_1^{\alpha_1}$ with $q = p_2^{\alpha_2} \dots p_k^{\alpha_k}$.

Evidently $(q, p_1) = 1$.

Now the positive factors of n are precisely of the types $d, dp_1, dp_1^2, \dots, dp_1^{\alpha_1}$ where d ranges over the divisors of q .

$$\begin{aligned}
\text{Thus } \sum_{d|n} \phi(d) &= \sum_{d|q} \phi(d) + \sum_{d|q} \phi(dp_1) + \dots + \sum_{d|q} \phi(dp_1^{\alpha_1}) \\
&= \sum_{d|q} \phi(d) + \sum_{d|q} \phi(d) \phi(p_1) + \dots + \sum_{d|q} \phi(d) \phi(p_1^{\alpha_1}) \\
&= \sum_{d|q} \phi(d) + \phi(p_1) \sum_{d|q} \phi(d) + \dots + \phi(p_1^{\alpha_1}) \sum_{d|q} \phi(d)
\end{aligned}$$

$$\begin{aligned}
&= \{1 + \phi(p_2) + \dots + \phi(p_1^{\alpha_1})\} \sum_{d|q} \phi(d) \\
&= p_1^{\alpha_1} q, \quad \text{by induction hypothesis} \\
&= n. \blacksquare
\end{aligned}$$

WORKED-OUT EXERCISES

Exercise 1. If a, b, c are natural numbers, then

$$\text{Max}\{\min(a, b), \min(a, c)\} = \text{Min}\{a, \text{max}(b, c)\}.$$

Solution. If a is the least among a, b, c , then $\text{LHS} = \text{Max}(a, a) = a$. Now $a \leq b$ and $a \leq c$ implies $a \leq \text{max}(b, c)$. So $\text{RHS} = a$. Hence the result holds in this case.

If a is the greatest among a, b, c , then $\text{LHS} = \text{Max}(b, c)$. Again $a \geq b, a \geq c$ implies $a \geq \text{max}(b, c)$. So $\text{Min}\{a, \text{max}(b, c)\} = \text{Max}(b, c)$. Hence the result holds in this case too.

So only two cases are to be discussed $b \leq a \leq c$ and $c \leq a \leq b$. Because of symmetry in b and c of the result to be proved, it is sufficient to discuss the case $b \leq a \leq c$.

$$\text{LHS} = \text{Max}(b, a) = a.$$

$$\text{RHS} = \text{Min}(a, c) = a.$$

Hence the result follows for all $a, b, c \in \mathbf{N}$.

Exercise 2. If $a, b, c \in \mathbf{N}$, then $[(a, b), (a, c)] = (a, [b, c])$.

In other words 'operation of taking HCF' is distributive over 'operation of taking LCM'.

Solution. If $a=1$, $\text{LHS} = [1, 1] = 1$ and $\text{RHS} = 1$.

If $b=1$, $\text{LHS} = [1, (a, c)] = (a, c)$ and $\text{RHS} = (a, c)$.

The case $c=1$ is similar to that of $b=1$.

So suppose each of a, b, c is greater than 1. We can write

$$a = p_1^{\alpha_1} p_2^{\alpha_2} \dots p_s^{\alpha_s}$$

$$b = p_1^{\beta_1} p_2^{\beta_2} \dots p_s^{\beta_s}$$

$$c = p_1^{\gamma_1} p_2^{\gamma_2} \dots p_s^{\gamma_s}$$

where for all $1 \leq i \leq s$, $\alpha_i, \beta_i, \gamma_i$ are non-negative integers. Let $\delta_i = \min(\alpha_i, \beta_i)$, $\theta_i = \min(\alpha_i, \gamma_i)$, $h_i = \max(\delta_i, \theta_i)$, $\phi_i = \max(\beta_i, \gamma_i)$ and $\psi_i = \min(\alpha_i, \phi_i)$.

As $(a, b) = p_1^{\delta_1} p_2^{\delta_2} \dots p_s^{\delta_s}$ (prove!) $(a, c) = p_1^{\theta_1} p_2^{\theta_2} \dots p_s^{\theta_s}$,

$$[b, c] = p_1^{\phi_1} p_2^{\phi_2} \dots p_s^{\phi_s}, \quad (a, [b, c]) = p_1^{\psi_1} \dots p_s^{\psi_s}$$

and $[(a, b), (a, c)] = p_1^{h_1} p_2^{h_2} \dots p_s^{h_s}$

so we are required to prove $h_i = \psi_i$ for all i .

In other words we are to prove $\max[\min(\alpha_i, \beta_i), \min(\alpha_i, \gamma_i)] = \min[\alpha_i, \max(\beta_i, \gamma_i)]$ for all $1 \leq i \leq s$.

But this is what we have proved in Exercise 1.

Exercise 3. If a, b are positive integers such that $(a, b) = c$ and $[a, b] = d$, then $ab = cd$.

Solution. Since $(a, b) = c$, $a = kc$, $b = lc$, we claim that $(l, k) = 1$. If $(l, k) = d'$, then $d' \mid l$ and $d' \mid k$.

So $d'c \mid a$ and $d'c \mid b$; but by definition of HCF this implies $d'c \mid c$ or $c = d'cu$ for some u .

In other words $d'u = 1$ so $d' = 1$. Hence our claim is established.

Now consider lkc . Clearly $a \mid lkc$ and $b \mid lkc$. Let $a \mid x$ and $b \mid x$, then $x = aw = by \Rightarrow kcw = lcy \Rightarrow kw = ly$. But $(l, k) = 1$ so $l \mid w$. This gives that $w = lm$ or $x = alm = lkc m$. Hence $lkc \mid x$. As a consequence $lkc = [a, b]$. By uniqueness of LCM we get $d = lkc$ or $cd = (kc)(lc) = ab$.

Exercise 4. Show that if $m \geq 2$ then the sum of all positive integers which are less than m and co-prime to m is $\frac{1}{2}m \phi(m)$.

Solution. Let $1 \leq x < m$ such that $(x, m) = 1$.

Now if $(m-x, m) = d$, then

$$d \mid m \text{ and } d \mid (m-x) \Rightarrow d \mid x, \text{ so } d \mid 1 \Rightarrow d = 1.$$

This gives $(m-x, m) = 1$. Also $x < m \Rightarrow m-x \geq 1$ and $x \geq 1$ gives $m-x < m$.

Let $X = \{x \mid 1 \leq x < m, (x, m) = 1\}$
and $Y = \{m-x \mid 1 \leq x < m, (x, m) = 1\}$.

There is 1-1 correspondence between X and Y given by $f(x) = m-x$. So $|X| = |Y| = \phi(m)$.

Thus $[x_1 + (m-x_1)] + [x_2 + (m-x_2)] + \dots + [x_n + (m-x_n)]$
 $= 2$ (sum of positive integers less than m and co-prime to m)
 $\underbrace{m + m + \dots + m}_{\phi(m) \text{ terms}} = 2S$, where S is required sum.

Hence
$$S = \frac{1}{2}m \phi(m).$$

PROBLEMS

1. If $a = p_1^{\alpha_1} \dots p_k^{\alpha_k}$ and $b = p_1^{\beta_1} p_2^{\beta_2} \dots p_k^{\beta_k}$ where the p_i 's are distinct prime numbers and where each $\alpha_i \geq 0$, $\beta_i \geq 0$ are integers, prove that

(i) $(a, b) = p_1^{\delta_1} p_2^{\delta_2} \dots p_k^{\delta_k}$, where $\delta_i = \min(\alpha_i, \beta_i)$ for each i .

(ii) $[a, b] = p_1^{\gamma_1} p_2^{\gamma_2} \dots p_k^{\gamma_k}$ where $\gamma_i = \max(\alpha_i, \beta_i)$ for each i .

2. Given a, b , on applying Euclidean algorithm successively we have: $a = q_0 b + r_1$, $0 \leq r_1 < |b|$, $b = q_1 r_1 + r_2$, $0 \leq r_2 < r_1$, $r_1 = q_2 r_2 + r_3$, $0 \leq r_3 < r_2$, $\dots$, $r_k = q_{k+1} r_{k+1} + r_{k+2}$, $0 \leq r_{k+2} < r_{k+1}$. Since the integers r_k are decreasing and are all non-negative, there

is a first integer n such that $r_{n+1}=0$. Prove that $r_n=(a, b)$. (We consider here $r_0 = |b|$).

3. If $c > 0$, $c \mid a$, $c \mid b$, $\left(\frac{a}{c}, \frac{b}{c}\right) = 1$ then $c=(a, b)$. Prove!
4. If $a, b, c \in \mathbf{N}$, then show that $([a, b], [a, c])=[a, (b, c)]$.
5. Show that if a and b are positive integers such that $(a, b)=1$ then $(a+b, a-b)$ is either 1 or 2.
6. Prove that $n > 1$ is a prime number if and only if for any a either $(a, n)=1$ or $n \mid a$.
7. If $a, b, c \in \mathbf{N}$ are such that $a^2+b^2=c^2$ then show that $a=2kmn$, $b=k(m^2-n^2)$, $c=k(m^2+n^2)$ for some $k, m, n \in \mathbf{N}$ such that $m > n$ and m, n are of opposite parity (i.e. either m is odd and n is even or m is even and n is odd).

Give an example to show that if we drop "identity" from the above question, the result need not be true.

[Hint: Consider $G = \{0, 2\}$ under multiplication modulo 4.]

5. Let G be a semi-group with identity e with the property that given $a \in G$ there exists an $a' \in G$ such that either $aa' = e$ or $a'a = e$. Prove that G is a group.

6. Let G be a semi-group with exactly one idempotent e such that."

(i) Each element in G has at least one left inverse with respect to e .

(ii) Each element in G has at most one right inverse with respect to e .

Show that G is a group.

7. If G is a group such that $(ab)^2 = a^2b^2$ for all $a, b \in G$ show that G must be Abelian.

Give example to show that this result does not hold for a semi-group.

8. If G is a group in which $(ab)^i = a^ib^i$ for three consecutive integers i and for all $a, b \in G$. Show that G is Abelian. Give an example to show that this result does not hold for semi-groups.

[Hint: Consider Example 14.]

9. Show that the conclusion of problem 8 does not follow if we assume the relation $(ab)^n = a^n b^n$ for just two consecutive integers.

[Hint: Consider $T = \{\pm 1, \pm i, \pm j, \pm k\}$, define $i^2 = j^2 = k^2 = -1$ and $ij = -ji = k, jk = -kj = i$ and $ki = -ik = j$. Then T is a group. T is called *Quaternion Group*. Take $n = 4, 5$.]

10. Show that if in a group G , $a^2 = e \forall a \in G$ then G is Abelian.

[Hint: $(ab)^2 = e \Rightarrow ab = (ab)^{-1} = b^{-1}a^{-1} = ba$.]

11. Show that if G is a finite semi-group with cross-cancellation laws, i.e. $xy = yz \Rightarrow x = z$, then G is an Abelian group.

[Hint: Consider $(xy)x = x(yx)$.]

12. If in a group G , there is no element c different from identity such that $c^2 = e$ and $(ab)^3 = (ba)^3$ holds for all a, b ; then show that G is commutative.

13. For any a in a group G and any integer n define a^n as follows:

$$a^0 = e$$

If n is positive, then

$$a^n = a^{n-1} \cdot a$$

If n is negative, say, $n = -m$ for some positive integer m then

$$a^n = (a^{-1})^m$$

Prove the following laws:

$$(i) (a^n)^{-1} = (a^{-1})^n = a^{-n}.$$

$$(ii) a^n a^m = a^{n+m}.$$

$$(iii) a^{nm} = (a^n)^m.$$

$$(iv) (x^{-1}ax)^n = x^{-1}a^nx \text{ for all } x \in G.$$

$$(v) \text{ If } ab = ba \text{ then } (ab)^n = a^n b^n.$$

14. If in a group G , $xy^n = y^{n+1}x$ and $yx^n = x^{n+1}y$ for some $n \in \mathbf{N}$, show that $x = y = e$ where e is the identity of G .

15. (Hayes) G is a semi-group. For every $x \in G$ there exists unique y in G such that $xyx = x$. Prove that G is a group.

16. Let S be a semi-group in which for some fixed integer $k \geq 1$, $x^{k+1} = x$ and $xy^kx = yx^ky$ for all x, y in S . Show that S is commutative.

17. Let S be a semi-group such that there exists an $e \in S$ satisfying $xe = x$ for all $x \in S$. Let, for each ordered pair $(x, y) \in S$, $x \neq y$ there exists a unique $w \in S$ such that $xw = y$. Show that S is group.

18. Let $J = \left\{ \begin{pmatrix} x & y \\ x & y \end{pmatrix} \mid x, y \in \mathbf{R} \text{ with } x+y \neq 0 \right\}$. Show that J is a

semi-group under matrix multiplication. Further show that J has a left identity and each element of J has a right inverse. Is J a group? (Ans. No)

19. Let S be a semi-group. If for all $x, y \in S$, $x^2y = y = yx^2$, prove that S is an Abelian group.

20. Give an example to show that the result of problem 19 above, may not be true if we have only $yx^2 = y$ for all $x, y \in S$ in place of $x^2y = y = yx^2 \forall x, y \in S$.

[Hint: Take $S = \{a, b\}$ and define $a^2 = a$, $ab = a$, $ba = b$ and $b^2 = b$.]

3. Subgroups

Let H be a non-void set and $*$ be a binary composition on H ; let A be a non-void subset of H . We recall that A is said to be *closed subset* of H with respect to the binary composition $*$, if for each pair $a, b \in A$, $a * b \in A$. In that case the mapping $f: A \times A \rightarrow A$ in which $f(a, b) = a * b$ is called a *binary composition on A induced by $*$* .

Let G be a group and let A be a non-void subset of G . It may happen that A is closed with respect to the binary composition on G and A itself is a group under the induced composition. Such a situation occurs quite often. This leads to the following concept.

Definition 2.14. *Subgroup.*

Let G be a group under a binary composition $*$ and let H be a non-void subset of G . Then H is called a *subgroup* of G if it satisfies the following:

(1) H is closed with respect to $*$. (Closure Property).

2

GROUPS

Introduction. In the previous chapter we studied the concept of relations, mappings and binary compositions. In this chapter we shall confine ourselves to the study of an algebraic system called Group. Before we formally define a group, let us emphasize that a group is a system consisting of a non-void set G and a binary composition on G satisfying some postulates. The importance of the concept of a group stems from the fact that a large number of known algebraic systems form groups; this will be illustrated by examples following the definition of a group. We wish to make the remark that the theory of groups has played very significant part in the development of Geometry and Number Theory, in recent times.

1. Definition and Examples

Definition 2.1. Group : A system $\langle G, * \rangle$, where G is a non-void set and $*$ is a binary composition on G , is called a group if it satisfies the following postulates:

- G.1. *Associative Law :* $a * (b * c) = (a * b) * c \quad \forall a, b, c \in G$.
- G.2. *Existence of Identity :* There exists an element $e \in G$ called an identity, such that $a * e = a = e * a \quad \forall a \in G$.
- G.3. *Existence of Inverse :* For each $a \in G$ there exists an element $a^{-1} \in G$, called an inverse of a , such that $a * a^{-1} = e = a^{-1} * a$.

If in addition to the above three postulates, the following postulate is also satisfied, the group is called a *Commutative* or an *Abelian group*.

- G.4. *Commutative Law :* $a * b = b * a \quad \forall a, b \in G$.

A group $\langle G, * \rangle$ for which the postulate G.4 does not hold is called a *Non-abelian group*. If the set G is finite then $\langle G, * \rangle$ is called a *finite group*, otherwise it is called an *Infinite group*.

For some definite purposes the simpler concept of a semi-group is important. So we define it.

Definition 2.2. A system $\langle G, * \rangle$ consisting of a non-void set G and a binary composition $*$ on G is called a *Semi-group* if it satisfies the following postulate:

S.G. 1. Associative Law: $a * (b * c) = (a * b) * c \quad \forall a, b, c \in G$.

In analogy with groups, we can talk of finite semi-groups, infinite semi-groups, Abelian and non-Abelian semi-groups. One notices that a group is always a semi-group while the converse is not true in general as will be shown later by examples.

In describing a particular group one has to specify the binary composition as well the set G on which it is defined. It was shown in the previous chapter that it is possible to define more than one binary composition on G . Thus the same set may be the *set part* of many different groups. Nevertheless if in a group $\langle G, * \rangle$ there is no likelihood of any confusion regarding the binary composition, we shall simply refer to the set G as a group. If we want to be particular about the composition, we shall say that the set G is a group under the binary composition $*$.

Examples from Numbers

EXAMPLE 1. As stated in Chapter 1, Section 3, usual addition $+$ on $\mathbf{Z}$ is commutative, associative, 0 is an identity for addition and for each $a \in \mathbf{Z}$, $-a$ is an additive inverse of a . Hence $\langle \mathbf{Z}, + \rangle$ is an Abelian group. Further $\langle \mathbf{Z}, + \rangle$ is an infinite group since $\mathbf{Z}$ is an infinite set.

EXAMPLE 2. Let $\mathbf{Z}'$ be the set of all non-negative integers. Since the sum of two non-negative integers is non-negative, $\mathbf{Z}'$ is closed under addition '+'. Addition on $\mathbf{Z}'$ is commutative, associative and 0 is an identity for $\mathbf{Z}'$. Thus $\langle \mathbf{Z}', + \rangle$ is an Abelian semi-group with identity. It is infinite as $\mathbf{Z}'$ is an infinite set. Now for any $a (\neq 0)$ in $\mathbf{Z}'$ there exists no non-negative integer b (i.e. there is no b in $\mathbf{Z}'$) such that $a + b = 0$. Thus $\mathbf{Z}'$ is a semi-group which is not a group.

EXAMPLE 3. Let $\mathbf{Q}^*$ be the set of all non-zero rational numbers. As the product of two non-zero rational numbers is a non-zero rational number, $\mathbf{Q}^*$ is closed under multiplication. For any $r, s, t \in \mathbf{Q}^*$ we know that (i) $r(st) = (rs)t$, (ii) $1 \in \mathbf{Q}^*$ and $1r = r1 = r$, (iii) $\frac{1}{r} \in \mathbf{Q}^*$ and $r \cdot \frac{1}{r} = 1$ (iv) $rs = sr$.

Thus $\mathbf{Q}^*$ is an Abelian group under the multiplication composition. $\mathbf{Q}^*$ is an infinite group.

EXAMPLE 4. On the similar lines as in Example 3, one can show that the set $\mathbf{R}^*, \mathbf{C}^*$ of non-zero real numbers and non-zero complex

numbers respectively form Abelian groups under multiplications compositions. Both are infinite groups.

EXAMPLE 5. Let n be a positive integer and G be the set of those complex numbers which are n th roots of unity. As there are n n th roots of unity, G has n elements. G is a finite Abelian group under multiplication composition.

EXAMPLE 6. Let $S = \{0, 1, 2, \dots, n-1\}$ where n is a positive integer. For $a, b \in S$ define $a * b$ to be the least non-negative integer c obtained as remainder when $a+b$ is divided by n (e.g. if $S = \{0, 1, 2, 3, 4\}$ then $2 * 4 = 1, 3 * 2 = 0$ etc.). This $*$ is a binary composition on S and is called *addition modulo n* . $a * b = r$ implies $a+b-r$ is divisible by n . We show that S is a group under this binary composition. Let $a, b, c \in S$.

(i) *Associativity.* Let $a * b = d$ and $(a * b) * c = d * c = r$. Since d is the least non-negative remainder got when $a+b$ is divided by n we have $a+b = d + kn$ for some integer k . Similarly $d+c = r + k_1n$ for some integer k_1 .

Thus $(a+b)+c = d+kn+c = r+(k_1+k)n$, $0 \leq r < n$. The last equation shows that r is the least non-negative remainder got on dividing $a+(b+c)$ by n . Similarly if $a * (b * c) = s$, then s is the least non-negative remainder got on dividing $a+(b+c)$ by n . However $(a+b)+c = a+(b+c)$. So the least non-negative remainders r, s must be equal.

Hence $(a * b) * c = a * (b * c)$.

(ii) *Existence for Identity.* For any $a \in S$, we claim that $a * 0 = a$. As $0 \leq a < n$, $0+a = a+0 = 0n+a$ gives a is the least non-negative remainder got on dividing a by n . So that $a * 0 = 0 * a = a$. Hence 0 is the identity of S .

(iii) *Existence of Inverse.* For any $a \in S$ if $a=0$ then $0 * 0 = 0$. If $a \neq 0$ then as $0 < n-a < n$, $n-a \in S$. Further $(n-a)+a = a+(n-a) = 1n+0$ gives $(n-a) * a = a * (n-a) = 0$. So that $n-a$ is the inverse of a under $*$. Hence S is a group under $*$. It can be verified that $\langle S, * \rangle$ is an Abelian group.

NOTATION. By $a+b \equiv c \pmod{n}$, $a, b, c \in \mathbb{Z}$ and n is a positive integer, we mean that n divides $(a+b-c)$.

EXAMPLE 7. Let $S = \{x \in \mathbb{Z} \mid 1 \leq x \leq n, \text{ and } (x, n) = 1\}$. Define $*$ on S as under:

For all $a, b \in S$, $a * b$ is the least positive integer obtained as remainder when ab is divided by n . This is called *multiplication modulo n* .

Firstly, we show that $*$ is a binary composition on S . Let $a * b = c$. Now c can't be zero otherwise $n \mid ab$ which is absurd as $(a, n) = 1$

and $(b, n)=1$, so $1 \leq c < n$. Further if $(c, n) \neq 1$ then there exists some prime number p such that $p \mid c$ and $p \mid n$. This gives $p \mid ab$, as $a * b = c$ implies $ab = c + kp$ for some integer k ; p being a prime either $p \mid a$ or $p \mid b$. Thus either p divides HCF of a and n or HCF of b and n which is impossible as both $(a, n)=1$ and $(b, n)=1$. Thus $(c, n)=1$. Hence $c \in S$. This proves our assertion that $*$ is a binary composition on S . This composition is called *multiplication modulo n* .

(i) *Associativity*. If $a * b = r_1$ and $(a * b) * c = r_1 * c = r_2$ so we get $r_2 * c = r$ i.e. $r_1 c = r_2 + k_2 n$ for some integer k_2 .

Now $ab = r_1 + k_1 n$ for some integer k_1 and $r_1 c = r_2 + k_2 n$

$$\Rightarrow (ab - k_1 n) c = r_2 + k_2 n$$

$$\Rightarrow (ab) c = r_2 + (k_2 + k_1 c) n.$$

This implies r_2 is the least non-negative remainder got on dividing $(ab)c$ by n . Similarly if $a * (b * c) = r_3$ then r_3 is the least non-negative integer obtained as remainder when $a(bc)$ is divided by n . But $(ab)c = a(bc)$, so $r_2 = r_3$.

Hence $(a * b) * c = a * (b * c)$.

(ii) *Existence of Identity*. Clearly $1 \in S$ and $\forall a \in S, 1 * a = a * 1 = a$ so 1 is identity of S .

(iii) *Existence of Inverse*. Let $a \in S$ then $(a, n)=1$.

So there exist integers x and y such that

$$ax + ny = 1. \text{ If } 1 \leq x < n \text{ and } (x, n)=1, x \in S.$$

If not, then by division algorithm in integers there exist integers q and r such that $x = qn + r, 0 \leq r < n$

Now $ax + ny = 1$

$$\Rightarrow aqn + ar + ny = 1$$

$$\Rightarrow ar = 1 + (-aq - y)(n)$$

So $a * r = 1$. Similarly $r * a = 1$. Now if $(r, n) \neq 1$, let p be a prime number dividing r and n . Then p will divide x so p divides 1 as $ax + ny = 1$, which is absurd.

Therefore $(r, n)=1$, hence $r \in S$.

Thus $\langle S, * \rangle$ is a group. It can be checked that $\langle S, * \rangle$ is Abelian.

NOTATION. By $ab = c \pmod{n}$ we shall mean $ab - c$ is a multiple of n .

REMARKS 1. One might have noted that in proving

$$(a * b) * c = a * (b * c)$$

we did not use a, b, c to be positive integers less than and relatively prime to n . Hence ' $*$ ' is associative in general.

2. As a particular case if we take $n=p$, a prime number then all integers $1, 2, \dots, p-1$ are relatively prime to p and are less than p . In this case $\langle S, * \rangle$ is called *group of integers under multiplication*.

modulo p .

Example from Vectors

EXAMPLE 8. The set V of all vectors in a plane starting from O (the origin) is an infinite Abelian group under vector addition.

In this group the null vector 0 is an additive identity and for each vector a , $-a$ is an additive inverse of a .

Examples from Matrices

EXAMPLE 9. For any positive integer n , the set C_n of all $n \times n$ matrices over the complex numbers form an infinite Abelian group under matrix addition as binary composition. The zero matrix is an additive identity and for any $A \in C_n$, $-A$ is an additive inverse of A .

EXAMPLE 10. Let G be the set of the following matrices

$$\begin{aligned} &\pm \begin{Bmatrix} 1 & 0 \\ 0 & 1 \end{Bmatrix}, \pm \begin{Bmatrix} i & 0 \\ 0 & -i \end{Bmatrix}, \\ &\pm \begin{Bmatrix} 0 & i \\ i & 0 \end{Bmatrix}, \pm \begin{Bmatrix} 0 & 1 \\ -1 & 0 \end{Bmatrix} \end{aligned}$$

where $i = \sqrt{-1}$. It can be easily verified that all the matrices are non-singular and set G is closed under the operation of matrix multiplication and G is a group under this operation. G is a finite non-commutative group (verify !)

Example of a group of Transformations

EXAMPLE 11. Let S be a non-void set and let $A(S)$ be the set of all permutations of S . We proved in Theorem 1.17 that for any $f, g \in A(S)$, fog denotes the mapping of S into S defined as $(fog)(s) = f[g(s)] \forall s \in S$, then $A(S)$ has the following properties:

For all $f, g, h \in A(S)$

- (i) $fog \in A(S)$.
- (ii) $fo(goh) = (fog)oh$.
- (iii) There exists $i \in A(S)$ such that $foi = f = iof$.
- (iv) For each $f \in A(S)$ there exists $f' \in A(S)$ such that $fof' = i = f'of$.

Thus the operation 'o' is a binary composition on $A(S)$ and $A(S)$ becomes a group under this operation. If S has n elements then $A(S)$ has $n!$ elements and $A(S)$ is a finite group. Further if S has 3 or

more elements than we found in Example 23 (Chapter 1), two $f, g \in A(S)$ such that $fog \neq gof$. Consequently $A(S)$ is a non-Abelian group whenever S contains more than two elements. If S is an infinite set then $A(S)$ is an infinite non-Abelian group.

PROBLEMS

1. Let Q^+ be the set of positive rational numbers.

Define $*$ on Q^+ as under

$$\text{For } a, b \in Q^+, a * b = \frac{ab}{3}.$$

Verify that $\langle Q^+, * \rangle$ is an Abelian group.

2. Let S be the set of all real numbers except -1 . Define $*$ on S by $a * b = a + b + ab \forall a, b \in S$. Show that $\langle S, * \rangle$ is an Abelian group.

3. Give an example of a non-Abelian group other than that given in the text above.

4. Let R^* be the set of all real numbers except 0. Define $*$ on R^* by $a * b = |a|b$ where $|a|$ denotes the absolute value of a .

Show that (i) $*$ is associative on R^* .

(ii) There exists a left identity for $*$.

and a right inverse for each element in R^* . Is $\langle R^*, * \rangle$ a group? Give reasons in support of your answer.

[Hint: 1, -1 are both left identities. Refer to § 2]

5. Give examples of systems $\langle G, * \rangle$, where G is a non-void set and $*$, a binary composition on G such that

(i) $\langle G, * \rangle$ satisfies G-1, G-2 but not G-3.

(ii) $\langle G, * \rangle$ satisfies G-1, G-3 but not G-2.

(iii) $\langle G, * \rangle$ satisfies G-2, G-3 but not G-1.

Hint: (iii) Consider $G = \{e, a, b\}$ and define $*$ on G with the help of the following table

$*$	e	a	b
e	e	a	b
a	a	e	a
b	b	b	e

[In this example $*$ is non-commutative. Can you give a similar example in which $*$ is commutative?]

6. Determine which of the following systems are groups. Give reasons why the remaining systems are not groups.

(i) The set G of all non-singular $n \times n$ matrices over complex numbers under matrix multiplication.

(ii) The set $\mathbf{N}$ under addition.

(iii) The set $\mathbf{R}$ under multiplication.

(iv) The set $G = \{1, a, b, c\}$ under a binary composition defined by the following composition table:

	1	a	b	c
1	1	a	b	c
a	a	1	c	b
b	b	c	1	a
c	c	b	a	1

[This is known as *Klein's four group*]

(v) The set $G = \{\alpha, \beta, \gamma, \delta\}$ under a binary composition defined by the following composition table:

	α	β	γ	δ
α	α	β	γ	δ
β	α	β	γ	δ
γ	α	β	γ	δ
δ	α	β	γ	δ

(vi) The set G of all n th roots of unity where n varies from 1 to ∞ , under multiplication.

(vii) The set $G = \{(m, n) \mid m, n \text{ integers such that either } m \neq 0 \text{ or } n \neq 0\}$ under the binary composition $*$ defined on G by

$$(a, b) * (c, d) = (ac - bd, ad + bc).$$

(viii) The set E of all even integers under addition.

(ix) The set E of all even integers under the binary compositions $*$ defined by

$$a * b = 2a + 2b$$

(x) The set K of all rational number of type $\frac{p}{q}$ where q is an odd integer, under usual addition of rational numbers.

7. In Problem (6) determine, if any, (i) all Abelian groups, (ii) all non-Abelian groups, (iii) all semi-groups which are not groups, and (iv) all the systems which are not semi-groups.

8. If the set of $n \times n$ matrices over complex numbers form a group under matrix multiplication, then show that either all the members of the set are non-singular matrices or all of them are singular.

2. Preliminaries

In this section we discuss the elementary properties of a group which are in fact the building blocks for the development of the fascinating and deep theory of groups. Let us recall our convention namely when we say that G is a group under a binary composition denoted multiplicatively or $\langle G, \cdot \rangle$ is a group. As per convention we shall use ab in place of $a \cdot b$.

Theorem 2.3. If G is a group then

- (1) *Identity element of G is unique.*
- (2) *Every $a \in G$ has unique inverse in G .*
- (3) *For every $a \in G$, $(a^{-1})^{-1} = a$.*
- (4) *For all $a, b \in G$, $(ab)^{-1} = b^{-1}a^{-1}$.*

REMARK. Before we proceed with the proof of the theorem let us try to understand as to what actually we want to prove and what is its significance. Let us recall that an element $e \in G$ is called an identity of G if $ae = a = ea \forall a \in G$. The postulate $G=2$ in the definition of G only assures the existence of an identity e , it does not assert that there is only one identity of G . So one may expect that there might exist another element $f \in G$ such that $af = a = fa \forall a \in G$. Part 1 of this theorem asserts that such a case is not possible. Similarly the postulate $G=3$ in the definition of group states that given $a \in G$ there exists at least one element $a^{-1} \in G$ such that $a^{-1}a = e = aa^{-1}$ where e is an identity of G but it does not state that there exists only one such element. In part 2 we prove that every element has a unique inverse. After following first and second parts of the theorem it becomes fairly easy to catch up the idea behind our purpose to prove third and fourth parts of above theorem.

Proof: (1) Let there exist two identities e and f of G . Then $ef = f$ as e is an identity and also $ef = e$ as f is an identity. Hence $ef = f = e$. This proves the uniqueness of identity.

(2) Let an element $a \in G$ have two inverses, say, x and y in G . So $x a = e = a x$ and $y a = e = a y$. This gives $x = x e = x (a y) = (x a) y = e y = y$.

Hence inverse of a is unique.

(3) Since for any $a \in G$, $a a^{-1} = e = a^{-1} a$, it follows from definition of inverse that $(a^{-1})^{-1} = a$.

$$(4) \text{ Now } (a b) (b^{-1} a^{-1}) = [a (b b^{-1})] a^{-1} \\ = (a e) a^{-1} = a a^{-1} = e.$$

$$\text{Also } (b^{-1} a^{-1}) (a b) = b^{-1} [a^{-1} (a b)] \\ = b^{-1} [(a^{-1} a) b] \\ = b^{-1} (e b) = b^{-1} b = e.$$

$$\text{Consequently } (a b)^{-1} = b^{-1} a^{-1}. \blacksquare$$

Theorem 2.4. (Cancellation Laws)

Let G be a group then for all $a, b, c \in G$

$$a b = a c \Rightarrow b = c \text{ (left cancellation law)}$$

$$b a = c a \Rightarrow b = c \text{ (right cancellation law)}$$

$$\text{Proof: } a b = a c \Rightarrow a^{-1} (a b) = a^{-1} (a c) \\ \Rightarrow (a^{-1} a) b = (a^{-1} a) c \\ \Rightarrow e b = e c \\ \Rightarrow b = c.$$

$$\text{Similarly } b a = c a \Rightarrow (b a) a^{-1} = (c a) a^{-1} \\ \Rightarrow b = c. \blacksquare$$

REMARK 1. It must be emphasized that the common factor in two equal products in a group can be cancelled only when it occurs on the same side. In fact for three elements a, b, c in a group G $a.b=c.a$ may not imply $b=c$ in general. To show this consider $S=\{1, 2, 3\}$ and the group $A(S)$. Define $f: S \rightarrow S, g: S \rightarrow S$ such that $f(1)=2, f(2)=3, f(3)=1$; $g(1)=2, g(2)=1, g(3)=3$. Clearly $f, g \in A(S)$. We leave it to the reader to verify that $fog=go$ (fog) but $f \neq fof$.

REMARK 2. In a semi-group the cancellation laws may not hold as is shown by the following example.

EXAMPLE 12. The set S of all 2×2 matrices over integers is a semi-group under multiplication. Now $A = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$, $B = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$ and $C = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$ are three elements of S such that $AB=AC$ but $B \neq C$.

REMARK 3. There are semi-groups which are not groups but they satisfy cancellation laws. This is shown as follows.

EXAMPLE 13. The set $\mathbb{N}$ is an Abelian semi-group under addition. It is well-known that for any three positive integers a, b, c ,

$a+b=a+c \Rightarrow b=c$. But this system is not a group.

Theorem 2.5. Any finite semi-group in which both cancellation laws hold, is a group.

Proof: Let S be a finite semi-group in which both cancellation laws hold.

Let $S = \{a_1, a_2, a_3, \dots, a_n\}$... (1)

Consider any $a_i \in S$, then

$a_1 a_i, a_2 a_i, a_3 a_i, \dots, a_n a_i$... (2)

are all distinct, since for any i and j ($1 \leq i, j \leq n$)

$a_i a_i = a_j a_i \Rightarrow a_i = a_j$ (right cancellation law).

Hence elements in (2) are all the elements in (1) in possibly some other order. So given any $a_i \in S \exists a_j \in S$ such that $a_i = a_j a_i$... (3)

In particular there exists some $a_k \in S$ such that $a_i = a_k a_i$.

Then $a_i a_i = a_i (a_k a_i) = (a_i a_k) a_i$

So by right cancellation law in S , $a_i = a_i a_k$ for all i .

Similarly considering the elements $a_i a_1, a_i a_2, \dots, a_i a_n$ and using left cancellation law we can find an element a_l such that $a_i a_i = a_i$ for all i .

Then we get $a_k = a_i a_k = a_i$

and $a_k a_i = a_i = a_i a_k$ for all $i=1, 2, \dots, n$.

Hence a_k is the identity element in S .

Let us write e for a_k . Taking $i=k$ in (3) we can say that there exists a_m such that $e = a_k = a_m a_i$. In a similar fashion, by considering the elements $a_i a_1, a_i a_2, \dots, a_i a_n$ we can find an element a_u such that $a_i a_u = e$.

Then $a_u = e a_u = (a_m a_i) a_u = a_m (a_i a_u) = a_m e = a_m$

and $a_m a_i = e = a_i a_m$. So $a_i^{-1} = a_m$.

Consequently every element in S has its inverse in S . S is such that for all $a, b, c \in S$

(i) $a(bc) = (ab)c$ (S is a semi-group).

(ii) There exists $e \in S$ such that

$ae = a = ea \forall a \in S$.

(iii) Given $a \in S, \exists a^{-1} \in S$ such that $aa^{-1} = e = a^{-1}a$.

Hence S is a group.

This proves the theorem. ■

Now if G is any finite group, then by Theorem 2.4, G satisfies cancellation laws. So we can say that G is a finite semi-group with cancellation laws. By taking this into consideration we have the following.

Theorem 2.6. A finite semi-group G is a group if and only if G satisfies both the cancellation laws ■

REMARK. Observe that in Example 13, $\langle \mathbb{N}, + \rangle$ is an infinite

semi-group with cancellation laws, but $\langle N, + \rangle$ is not a group. This means that preceding theorem cannot be generalized to infinite semi-groups.

Theorem 2.7. Given a, b in a group G , the equations $ax=b$ and $ya=b$ have unique solutions for x and y in G .

Proof: Since $a(a^{-1}b) = (aa^{-1})b = eb = b$, $x = a^{-1}b$ is a solution of $ax=b$ in G . Similarly $y = ba^{-1}$ is a solution of the equation $ya=b$ in G . The uniqueness of the solution follows from the cancellation laws in G . ■

Theorem 2.8. A non-void set G together with a binary composition denoted multiplicatively is a group if and only if

(1) $a(bc) = (ab)c$ for all $a, b, c \in G$.

(2) Given any $a, b \in G$ the equations $ax=b$ and $ya=b$ have solutions in G .

Proof: "Only if", part.

Let G be a group. By the definition of group, (1) holds in G and (2) follows from Theorem 2.7.

"If" part

Let G be a non-void set with a multiplicative binary composition satisfying (1) and (2). To show that G is a group we only need to show that G contains an identity and each element of G has an inverse.

(1) "Existence of Identity".

Consider any $a \in G$ then by (2) the equation $ax=a$ has solution in G . Thus there exists an element $e \in G$ such that $ae=a$. Now given $b \in G$, by (2) $\exists y \in G$ such that $b=ya$. Then

$$be = (ya)e = y(ae) = ya = b$$

So $be=b$ for all $b \in G$.

Similarly since the equation $ya=a$ has solution in $G \exists f \in G$ such that $fa=a$ and $fb=b \forall b \in G$. In particular $fe=f$ since $be=b$ for all $b \in G$. Also $fe=e$ since $fb=b \forall b \in G$. Consequently $e=f$ and we get that $be=b=eb \forall b \in G$.

Hence e is the identity of G .

(II) "Existence of Inverse".

Given $a \in G$ by (2) $\exists a', a'' \in G$ such that $aa' = e$, $a''a = e$.

Then $a' = ea' = (a''a)a' = a''(aa') = a''e = a''$ and we find that $a'a = e = aa'$.

Hence $a' = a^{-1}$ exists in G .

Thus from (I) and (II) it follows that G is a group. ■

This theorem gives a characterization of a group so we may alternately define a group as follows:

Definition 2.9. A system $\langle G, . \rangle$ where G is a non-void set and

'.' a multiplicatively denoted binary composition is called a *Group* if it satisfies the following postulates.

For all $a, b, c \in G$

$$(1) a(bc) = (ab)c$$

$$(2) \text{ The equations } ax=b \text{ and } ya=b$$

have solutions in G .

REMARK. However if G is a semi-group and if we are given that for all $a, b \in G$ only the equation $ax=b$ has a solution in G then G may not be a group. This is shown by the following example.

EXAMPLE 14. Let G be any set having at least two elements. Given $a, b \in G$ define $a.b=b$.

(I) Let $a, b, c \in G$. Then $a.(b.c)=a.c=c$ and $(a.b).c=b.c=c$. Hence $(a.b).c=a.(b.c)$.

(II) Since $a.b=b$, $x=b$ is a solution of $ax=b$. Consider two distinct elements a, b in G . Then $a.b=b$, $b.b=b$. Consequently $b.b=a.b$ but $a \neq b$. So G does not satisfy cancellation laws. Hence G is not a group.

Theorem 2.10. A non-void set G is a group under a binary composition denoted multiplicatively if and only if it satisfies the following:

$$(1) a(bc) = (ab)c \quad \forall a, b, c \in G.$$

$$(2) \text{ There exists an element } e \in G \text{ such that } ae=a \quad \forall a \in G.$$

$$(3) \text{ For each } a \in G, \exists a' \in G \text{ such that } aa'=e.$$

Proof.: The "only if" follows from the definition 2.1 of a group.

"If". Suppose G is a non-void set with binary composition satisfying (1), (2) and (3). Let $a \in G$, by (3), $\exists a' \in G$ such that $aa'=e$. Again by (3) $\exists a'' \in G$ such that $a' a''=e$.

$$\begin{aligned} \text{Now } a'a &= (a'a)e = (a'a)(a'a'') = a'[a(a'a'')] \\ &= a'[(aa')a''] \\ &= a'(ea'') = (a'e)a'' \\ &= a'a'' \\ &= e \end{aligned}$$

$$\text{Thus } aa'=e=a'a.$$

$$\text{Further } ea=(aa')a=a(a'a)=ae=a \text{ so } ae=a=ae.$$

Hence G is a group. ■

In same way we can prove the following.

Theorem 2.11. A non-void set G is a group under a binary composition on G denoted multiplicatively if and only if it satisfies the following:

$$(1) a(bc) = (ab)c \quad \forall a, b, c \in G.$$

$$(2) \exists e \in G \text{ such that } ea=a \quad \forall a \in G.$$

$$(3) \text{ For each } a \in G, \exists a' \in G \text{ such that } a'a=e. \blacksquare$$

Theorems 2.10 and 2.11 provide the following alternative defini-

tions of a group.

Definition 2.12. A system $\langle G, . \rangle$ where G is a non-void set and $'.'$ is a binary composition on G is called a *group* if it satisfies the following postulates:

- (1) $a.(b.c)=(a.b).c \quad \forall a, b, c \in G$
- (2) $\exists e \in G$ such that $a.e=a \quad \forall a \in G$
- (3) For each $a \in G \exists a' \in G$ such that $a.a'=e$.

Definition 2.13. A system $\langle G, . \rangle$ where G is a non-void set and $'.'$ is a binary composition on G is called a *group* if it satisfies the following postulates. For all $a, b, c \in G$

- (1) $a.(b.c)=(a.b).c$.
- (2) There exists an element $e \in G$ such that $e.a=a$.
- (3) For each $a \in G$, there exists $a' \in G$ such that $a'.a=e$.

Definition 2.12 is called the *right handed* definition of a group and definition 2.13 is called the *left handed* definition of a group. It should be noted that a semi-group G may be such that G contains a left identity e and each element $a \in G$ may possess a right inverse but G may fail to be a group. This is shown by the following example.

EXAMPLE 15. Let G be a semi-group considered in Example 14 and let e be any fixed element of G . Then by definition

- (I) $ea=a \quad \forall a \in G$. Hence e is a left identity.
- (II) For each $a \in G$, since $ae=e$ so that e is a right inverse of a with respect to e .

Thus G has left identity e and each element of G has a right inverse in G but we know that G is not a group.

We leave it to the readers to construct similarly an example of a semi-group G which has a right identity and left inverse for each of its elements but it fails to be a group.

WORKED-OUT EXERCISES

Exercise 1. If in a group G , $xy^2=y^3x$ and $yx^2=x^3y$, then show that $x=y=e$ where e is the identity of G .

$$\begin{aligned} \text{Solution. } xy^2 &= y^3x \Rightarrow x = y^3xy^{-2} \\ x^2 &= xy^3xy^{-2} = xy^2yxy^{-2} = y^3xyxy^{-2} \\ x^2y &= y^3xyxy^{-1} \end{aligned} \quad \dots (1)$$

$$\begin{aligned} \text{Now } yx^2 &= x^3y \text{ implies } yx^2 = xy^3xyxy^{-1} \\ \Rightarrow x^2 &= y^{-1}xy^3xyxy^{-1} \\ \Rightarrow x^2y &= y^{-1}xy^3xyx \end{aligned} \quad \dots (2)$$

By (1) and (2) we obtain

$$\begin{aligned} y^3xyxy^{-1} &= y^{-1}xy^3xyx \Rightarrow y^4xyx = xy^3xyxy \\ \Rightarrow y^4xyx &= xy^2yx \quad yxy = y^3xyxy \end{aligned}$$

$$\Rightarrow (yx)^2 = (xy)^3$$

Interchange x any y in (3) to get

... (3)

$$(xy)^2 = (yx)^3$$

Now (3) and (4) imply

... (4)

$$(xy)^2 = (yx)^3 = (yx)^2(yx) = (xy)^3(yx)$$

$$\Rightarrow e = xy^2x \Rightarrow x^{-2} = y^2$$

Further $xy^2 = y^3x \Rightarrow x x^{-2} = yx^{-2}x$

$$\Rightarrow x^{-1} = yx^{-1} \Rightarrow y = e.$$

Lastly $yx^2 = x^3y \Rightarrow ex^2 = x^3e \Rightarrow x = e.$

Exercise 2. (Isbell). G is a group and there exist two relatively prime positive integers m and n such that $a^m b^m = b^m a^m$ and $a^n b^n = b^n a^n$ for all $a, b \in G$. Prove that G is Abelian.

Solution. Since $(m, n) = 1$, we get $mx + ny = 1$ for some $x, y \in \mathbb{Z}$.

Now $(a^m b^n)^{mx} = a^m (b^n a^m)^{mx-1} b^n$

$$= a^m (b^n a^m)^{mx} (b^n a^m)^{-1} b^n$$

$$= (b^n a^m)^{mn} a^m a^{-m} b^{-n} a^n$$

$$= (b^n a^m)^{mx}.$$

... (1)

Similarly it can be proved that

$$(a^m b^n)^{ny} = (b^n a^m)^{ny}.$$

... (2)

From (1) and (2) we get

$$a^m b^n = (a^m b^n)^{mx+ny}$$

$$= (b^n a^m)^{nx+ny} = b^n a^m.$$

... (3)

Finally $ab = a^{mx+ny} b^{nx+ny}$

$$= a^{mx} (a^{ny} b^{nx}) b^{ny}$$

$$= a^{mx} b^{mx} a^{ny} b^{ny} \quad \text{by (3)}$$

$$= b^{mx} a^{mx} b^{ny} a^{ny} \quad \text{by hypothesis}$$

$$= b^{mx+ny} a^{mn+ny} \quad \text{by (3)}$$

$$= ba.$$

Hence G is Abelian.

Exercise 3. S is a finite semi-group. Show there is an element $e \in S$ such that $e^2 = e$.

Solution. Let $x \in S$, as S is finite $x, x^2, x^3, \dots$ all cannot be distinct. So there exist integers $m > n$ such that $x^m = x^n$. This gives $x^{n+k} = x^n$ where $k = m - n$.

Now $x^{2n+k} = x^n x^{n+k} = x^{2n}$. By principle of induction it can be proved that $x^{mn+k} = x^{mn}$ for any $m \in \mathbb{N}$.

Also $x^{mn+2k} = x^{mn+k} x^k = x^{mn} x^k = x^{mn+k} = x^{mn}$, $x^{mn+3k} = x^{mn+2k} x^k$
 $= x^{mn+k} = x^{mn}$ and so on.

Again by using principle of induction it can be proved that $x^{mn+lk} = x^{mk}$ for any $l \in \mathbb{N}$. In particular we get $x^{kn+nk} = x^{kn}$ i.e. $x^{2nk} = x^{nk}$.

Put $x^{nk} = e$.

Exercise 4. G is a group and $a \in G$. Define a binary composition $*$ on G such that $\langle G, * \rangle$ is a group with a as its identity.

Solution. We define $*$ on G as under

$$x * y = xa^{-1}y \text{ for all } x, y \in G.$$

1. $*$ is Associative. Let $x, y, z \in G$.

$$(x * y) * z = (xa^{-1}y) * z = (xa^{-1}y)a^{-1}z \\ = xa^{-1}ya^{-1}z.$$

$$\text{Again } x * (y * z) = x * (ya^{-1}z) = xa^{-1}(ya^{-1}z) \\ = xa^{-1}ya^{-1}z$$

$$\text{Hence } (x * y) * z = x * (y * z)$$

2. For any $x \in G$,

$$x * a = xa^{-1}a = xe = x \text{ where } e \text{ is identity of } G$$

$$\text{Similarly } a * x = aa^{-1}x = ex = x.$$

Hence a is identity of G

3. Let $x \in G$.

$$\text{Then } x * (ax^{-1}a) = xa^{-1}(ax^{-1}a) = x \cdot 1 \cdot a = ea = a..$$

$$\text{Further } (ax^{-1}a) * x = (ax^{-1}a) a^{-1}x = ax^{-1}x = ae = a.$$

$$\text{Hence } x^{-1} = ax^{-1}a.$$

Consequently $\langle G, * \rangle$ is a group with a as its identity.

Exercise 5. Show that the equation $x^2ax = a^{-1}$ is solvable for x in a group G if and only if a is the cube of some element in G .

Solution. Suppose $x^2ax = a^{-1}$ is solvable in G , there exists $c \in G$ such that $c^2ac = a^{-1}$. This gives $ca ca = c^{-1} \Rightarrow ca ca c = e \Rightarrow ca ca ca = ea = a \Rightarrow (ca)^3 = a$.

Conversely let $a = b^3$ for some b in G . Then $x = b^{-2}$ is a solution of $x^2ax = a^{-1}$, since, $x^2ax = b^{-2}b^3b^{-2} = b^{-3}$ while $a^{-1} = b^{-3}$.

Exercise 6. A semi-group S is called regular if for every $y \in S$ there exist $a \in S$ such that $ya = y$.

Let S be a semi-group with at least three elements and $x \in S$ is such that $S - \{x\}$ is a group. Prove that S is regular if and only if $x^2 = x$.

Solution. If $x^2 = x$, then $x^3 = x^2 = x$ i.e. $xxx = x$. Now $G = S - \{x\}$ is a group, let $y \in S - \{x\}$, then $yy^{-1}y = ey = y$ where e is identity of G . As $S = G \cup \{x\}$, S is regular.

Conversely let S be regular. We claim that for any $y \in G$, $xy = x = yx$. Suppose $xy \neq x$ then $xy \in G$ so $(xy)y^{-1} = x \in G$ which is absurd. Hence $xy = x$. Similarly $yx = x$. Since S is regular, there is an element $a \in S$ such that $xax = x$.

$$\text{If } a \in G \text{ then } ax = x \Rightarrow xx = x \Rightarrow x^2 = x.$$

$$\text{If } a \notin G, a = x. \text{ Then } xax = x \Rightarrow x^3 = x.$$

Let $x^2 \neq x$, then $u = x^2 \in G$. As S has at least three elements, there exists $v (\neq u)$ in G .

Now $zu=v$ is solvable in G . Because $z \in G$, $zu=zx^2=(zx)x=xx^2=x^3=x^2=u$, i.e. $u=v$; which is absurd hence $x^2=x$.

REMARK. If in a semi-group S , there exists an element $t \in S$ such that $xt=tx=t$ for all $x \in S$ then t is called a 'zero' of S . Thus in Exercise 6 above we note that if S is a regular semi-group with at least three elements and $x \in S$ such that $S-\{x\}$ is a group, then x is 'zero' of S .

PROBLEMS

1. *Definition.* In a semi-group G , an element a is called an *idempotent* if $a^2=a$.

Show that if G is a group then $a \in G$ is an idempotent if and only if $a=e$, the identity of G .

2. Give an example of a semi-group containing more than one idempotent but containing no identity.

3. (a) We know from Theorem 2.3 that a finite semi-group satisfying cancellation laws is a group. Give an example of a non-void set G with a binary composition such that G satisfies both the laws of cancellation but G is not a group.

[Hint: Consider set $G=\{1, \alpha, \beta, \gamma, \delta\}$ together with binary composition given by the following multiplication table. Since no column or row consists of same elements in two or more places hence cancellation laws hold.]

	1	α	β	γ	δ
1	1	α	β	γ	δ
α	α	1	γ	δ	β
β	β	δ	α	1	γ
γ	γ	β	δ	α	1
δ	δ	γ	1	β	α

(b) Show that any set G having 1, 2, 3 or 4 elements is a group under any binary composition which is cancellative and for which identity exists.

4. Prove that a finite semi-group G with identity is a group if and only if G contains only one idempotent.

(2) H is a group under the binary composition of H induced by $*$.

Let e be the identity of a group G . Then trivially G and $\{e\}$ are subgroups of G . These subgroups are called *Improper subgroups* of G . A subgroup H of a group G , is called a *Proper subgroup* if it is different from G as well as from $\{e\}$.

If G is a group and H is a subgroup of G , for the sake of convenience we shall usually denote the induced binary composition of H by the same symbol which denotes the binary composition on G .

Examples of Subgroups

EXAMPLE 16. The set E of all even integers is a subgroup of additive group $\mathbf{Z}$.

EXAMPLE 17. The set $\mathbf{Z}$ is a subgroup of the additive group $\mathbf{Q}$.

EXAMPLE 18. Let $G = \{1, -1, i, -i\}$, $H = \{1, -1\}$ where $i^2 = -1$.

Then G is a group under usual multiplication of complex numbers and H is a subgroup of G .

EXAMPLE 19. Let G be the multiplicative group of all non-singular 2×2 matrices over complex numbers. Let H be the set of the following eight matrices

$$\begin{aligned} \pm \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}, & \quad \pm \begin{bmatrix} i & 0 \\ 0 & -i \end{bmatrix}, \\ & \quad \pm \begin{bmatrix} 0 & i \\ i & 0 \end{bmatrix}, \quad \pm \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \end{aligned}$$

Clearly $H \subseteq G$. As was claimed in Example 10, H is a group under matrix multiplication. Hence H is a subgroup of G .

EXAMPLE 20. The set $\mathbf{Q}^+$ of all non-zero positive rational numbers is a subgroup of the multiplicative group $\mathbf{Q}^*$ of all non-zero rational numbers.

From the fact that a group has only one idempotent [Prob. 1 § 2], and the inverse of any element of G is unique, the following is immediate.

Lemma 2.15. If G is a group and H is a subgroup of G , then

(1) e , the identity of G , is also the identity of H ,

(2) for each $a \in H$, $a^{-1} \in H$. ■

We now prove a theorem which supplies us a criterion to check whether a given subset of a group is a subgroup or not.

Theorem 2.16. A non-void subset H of a group G is a subgroup of G if and only if $a, b \in H \Rightarrow ab^{-1} \in H$.

Proof: Let H be a subgroup of G and let $a, b \in H$. By lemma 2.15 (2), $b^{-1} \in H$. Consequently by the closure property of H , $ab^{-1} \in H$.

Conversely let H be a non-void subset of G such that for all $a, b \in H; ab^{-1} \in H$.

(I) Since H is non-void, there exists $a \in H$.

Thus $aa^{-1} \in H$ i.e., $e \in H$. So H contains identity.

(II) Let $b \in H$. As $e \in H, eb^{-1}$ i.e., $b^{-1} \in H$.

Hence $b \in H \Rightarrow b^{-1} \in H$.

(III) $a, b \in H \Rightarrow a$ and $b^{-1} \in H$ by (II)

$$\Rightarrow a(b^{-1})^{-1} \in H$$

i.e. $ab \in H$, since $(b^{-1})^{-1} = b$.

Consequently H is a closed subset of G .

Trivially the induced binary composition on H is associative because the binary composition on G is associative. Hence H is a subgroup of G . ■

Theorem 2.17. Let G be a group and H be a finite non-void subset of G . Then H is a subgroup of G if and only if $ab \in H$ for all $a, b \in H$.

Proof: If H is a subgroup of G then by definition

$$a, b \in H \Rightarrow ab \in H.$$

Conversely let H be a finite non-void subset of G such that

$$a, b \in H \Rightarrow ab \in H.$$

Thus H is finite semi-group. Further the cancellation laws hold in H since the same hold in G . Hence H is a subgroup (Theorem 2.6). ■

We know that given two integers $a, b, a \equiv b \pmod{m}$ if and only if $m \mid a-b$ or equivalently $a-b$ is an integral multiple of m . Thus $a \equiv b \pmod{m}$ if and only if $a-b \in (m)$ i.e., the subgroup of $\mathbb{Z}$ consisting of multiples of m .

Hence congruence between two integers a and b is closely related to the subgroup (m) .

Definition 2.18. Right congruence modulo a subgroup.

Let G be a group and H be a subgroup of G . Given $a, b \in G, a$ is said to be *right congruent to b modulo H* {symbolically $a \equiv_r b \pmod{H}$ } if and only if $ab^{-1} \in H$.

Theorem 2.19. If G is a group and H is a subgroup of G . Then the relation $\equiv_r$ of right congruence modulo H is an equivalence relation on G . Further for any $a \in G$ the set $\{ha \mid h \in H\}$ is the equivalence class to which a belongs.

Proof: Let $a, b \in G$ and e be identity of H .

(1) *Reflexivity.* Since $aa^{-1} = e \in H, a \equiv_r a \pmod{H}$.

(2) *Symmetry.* $a \equiv_r b \pmod{H} \Rightarrow ab^{-1} \in H$.

$$\Rightarrow (ab^{-1})^{-1} \in H$$

$$\text{i.e. } ba^{-1} \in H$$

$$\Rightarrow b \equiv_r a \pmod{H}.$$

- (3) *Transitivity.* $a \equiv_{\tau} b \pmod{H}$, $b \equiv_{\tau} c \pmod{H}$
 $\Rightarrow ab^{-1} \in H$, $bc^{-1} \in H$
 $\Rightarrow ac^{-1} = (ab^{-1})(bc^{-1}) \in H$
 $\Rightarrow a \equiv_{\tau} c \pmod{H}$

Thus the relation of right congruence modulo H is an equivalence relation on G . Let $Cl(a)$ denote the equivalence class to which a belongs i.e., $Cl(a) = \{b \in G \mid b \equiv_{\tau} a \pmod{H}\}$.

Let Ha denote the set $\{ha \mid h \in H\}$.

Now $b \in Cl(a) \Rightarrow b \equiv_{\tau} a \pmod{H}$

$$\Rightarrow ba^{-1} \in H \Rightarrow b = (ba^{-1})a \in Ha. \text{ Thus } Cl(a) \subseteq Ha.$$

Again $c \in Ha \Rightarrow c = ha$ for some $h \in H \Rightarrow ca^{-1} = h \in H$

$$\Rightarrow c \equiv_{\tau} a \pmod{H} \Rightarrow c \in Cl(a). \text{ Thus } Ha \subseteq Cl(a).$$

Hence $Cl(a) = Ha$. ■

Definition 2.20. *Right coset.* Let G be a group and H be a subgroup of G . For any $a \in G$, the set $Ha = \{ha \mid h \in H\}$ is called a *right coset* of H determined by a .

Thus we find that the right coset Ha is nothing else but the equivalence class, determined by the relation of right congruence modulo H , to which a belongs.

In analogy with theorem 2.19 and definitions 2.18 and 2.20 we have

Definition 2.21. *Left congruence modulo a subgroup:* Let G be a group and H be a subgroup of G . Given $a, b \in G$, a is said to be *left congruent to b modulo H* {symbolically $a \equiv_{\iota} b \pmod{H}$ } if and only if $a^{-1}b \in H$.

Theorem 2.22. *If G is a group and H is a subgroup of G . Then the relation $\equiv_{\iota}$ of left congruence modulo H is an equivalence relation on G . Further for any $a \in G$ the set $\{ah \mid h \in H\}$ is the equivalence class to which a belongs.* ■

Definition 2.23. *Left coset.* Let G be a group, and H be a subgroup of G . For any $a \in G$, the set $aH = \{ah \mid h \in H\}$ is called a *left coset* of H determined by a .

We now prove a very useful theorem on cosets without using the concepts of congruences modulo a subgroup. However this theorem is implicitly contained in Theorem 2.19.

Theorem 2.24. *Let G be a group, H be a subgroup of G and $a, b \in G$. Then*

- (1) $Ha = H \Leftrightarrow a \in H$.
- (2) $Ha = Hb \Leftrightarrow ab^{-1} \in H$.
- (3) $aH = bH \Leftrightarrow a^{-1}b \in H$.

Proof: (1) Now $Ha = H \Rightarrow ea \in Ha = H$ as $e \in H$
 $\Rightarrow a \in H$.

i.e. $Ha=H \Rightarrow a \in H$.

Let $a \in H$, then $ha \in H \forall h \in H$ as H is a subgroup, so by definition $Ha \subseteq H$.

Now for any $h \in H$, $h=(ha^{-1})a \in Ha$. This yields $H \subseteq Ha$.

Thus $Ha=H$.

Hence $Ha=H \Leftrightarrow a \in H$.

(2) $Ha=Hb \Rightarrow a \in Hb$ since $a=ea \in Ha$.

$$\Rightarrow a = hb \text{ for some } h \in H$$

$$\Rightarrow ab^{-1}=h \in H.$$

Again $ab^{-1} \in H \Rightarrow ab^{-1}=h$ for some $h \in H$

$$\Rightarrow a=hb$$

$$\Rightarrow Ha=H(hb)=(Hh)b$$

$$=Hb, \text{ since } Hh=H \text{ by (1)}$$

Hence $Ha=Hb \Leftrightarrow ab^{-1} \in H$.

The proof of (3) is similar to that of (2). ■

If a group G is Abelian and H is a subgroup of G , then for any $a \in G$, $h \in H$, $ah=ha$ and so $Ha=aH$. But if a group G is not Abelian we cannot say that for any $a \in G$ and any subgroup H of G , $aH=Ha$. However a group G may contain a subgroup H for which $Ha=aH$ holds for all $a \in G$, as is shown by the following example.

EXAMPLE 21. Let G be the group of matrices

$$\pm \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \pm \begin{bmatrix} i & 0 \\ 0 & -i \end{bmatrix}, \quad \pm \begin{bmatrix} 0 & i \\ i & 0 \end{bmatrix},$$

$$\pm \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$$

(Cf. Example 10) and let H be the subgroup consisting of matrices

$$\pm \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \pm \begin{bmatrix} i & 0 \\ 0 & -i \end{bmatrix}$$

Then H is a subgroup of G (verify!). We leave it to the readers to verify that $aH=Ha$ for all $a \in G$.

Definition 2.25. A subgroup H of a group G is said to be a *normal subgroup* of G if $Ha=aH$ for all $a \in G$.

For a group G , G and $\{e\}$ are always normal subgroups of G and these subgroups are called *trivial normal subgroups*. A group $G \neq \{e\}$ which does not have any non-trivial normal subgroup is called a *simple group*. In Example 21 above, H is a non-trivial normal subgroup of G .

A normal subgroup is also known as an *Invariant* or *self-conjugate* subgroup. Clearly every subgroup of an Abelian group is a normal subgroup. Following theorem gives an elegant and useful characterization of a normal subgroup.

Theorem 2.26. A subgroup H of a group G is normal if and only if $g^{-1}hg \in H$ for every $h \in H, g \in G$.

Proof: Let H be a normal subgroup of G . Let $h \in H, g \in G$

Then $Hg = gH$ (Definition of normal subgroup)

Now $hg \in Hg = gH$

so $hg = gh_1$ for some $h_1 \in H$.

i.e. $g^{-1}hg = h_1 \in H$.

Conversely let H be such that

$$g^{-1}hg \in H \quad \forall h \in H, g \in G.$$

Consider $a \in G$ For any $h \in H, a^{-1}ha \in H$

Therefore $ha = a(a^{-1}ha) \in aH$.

Consequently $Ha \subseteq aH$.

Let $b = a^{-1}$

then $b^{-1}hb \in H$

But $b^{-1}hb = (a^{-1})^{-1}ha^{-1} = aha^{-1}$

This gives $aha^{-1} \in H$

so that $ah = (aha^{-1})a \in Ha$

which proves that $aH \subseteq Ha$.

Hence $aH = Ha$. ■

This theorem shows that, equivalently a subgroup H of a group G can be defined to be a normal subgroup if

$$g^{-1}hg \in H \quad \forall h \in H, g \in G.$$

Definition 2.27. Order of a group. The number of elements in a group is called the order of the group.

The order of a group G will be denoted by $o(G)$. A group of finite order is called a *finite group*. By using the concept of cosets we prove a theorem due to Lagrange which expresses a relationship between the order of a finite group and the order of its subgroup. This theorem has played a fundamental role in the development of theory of finite groups.

Theorem 2.28. (Lagrange). The order of any subgroup of a finite group divides the order of the group.

Proof: Let G be a finite group and H a subgroup of G .

Suppose $o(H) = n$. For any $a \in G$, define $f: H \rightarrow Ha$ such that $f(h) = ha$.

This mapping is onto as each member of Ha is of type $ha, h \in H$. Further for any two elements $h_1, h_2 \in H, f(h_1) = f(h_2) \Rightarrow h_1a = h_2a \Rightarrow h_1 = h_2$ (right cancellation law).

Hence f is a 1-1 mapping of H onto Ha . This means that

$$o(H) = o(Ha).$$

Let $Ha_1, Ha_2, \dots, Ha_t$ be the totality of distinct right cosets of H in G . Then these t right cosets constitute the totality of distinct equivalence classes in G determined by the relation of right

congruence modulo H (Theorem 2.19). Thus $Ha_1, Ha_2, \dots, Ha_t$ are disjoint subsets of G and $G = Ha_1 \cup Ha_2 \cup \dots \cup Ha_t$.

As seen above $o(Ha_i) = o(H) = n \quad \forall i = 1, 2, \dots, t$.

Therefore $o(G) = nt$. Hence $o(H) \mid o(G)$. ■

It is evident from the proof of the above theorem that $\frac{o(G)}{o(H)}$ is equal to the number of right cosets of H in G . Similarly it can be proved that $\frac{o(G)}{o(H)}$ is equal to the number of left cosets of H in G .

Definition 2.29. Index of a subgroup:

Let G be a group and H , a subgroup of G . Then the number of right (left) cosets of H in G is called the *Index* of H in G .

The index of H in G is denoted by $[G : H]$.

Hence the theorem of Lagrange can be re-stated as follows:

Theorem 2.30. For any subgroup H of a finite group G

$$o(G) = o(H) [G : H]. \blacksquare$$

For a finite group G , if k is a positive integer such that $k \mid o(G)$, it is not necessary that G contains a subgroup of order k [see Theorem 3.30]. However if G is Abelian and k is a divisor of $o(G)$ then G must contain a subgroup of order k . This will be proved later.

Definition 2.31. Order of an element:

Let G be a group and let $a \in G$. Then a is said to be of *finite order* n , if n is the least positive integer such that $a^n = e$, the identity of G .

If for no positive integer k , $a^k = e$, then a is said to be of *infinite order*.

The symbol $o(a)$ shall denote the order of a .

Definition 2.32. Cyclic Group. A group G is said to be a cyclic group if there exists an element $b \in G$, such that every element of G is a power of b . b is called a *generator* of G and we denote G by $\langle b \rangle$.

If the composition in G were denoted additively then we could say that G is a cyclic group if there exists an element a of G such that every element of G is of the form na where n is an integer.

EXAMPLE 22. Let $G_n = \left\{ e^{\frac{2\pi i r}{n}} \mid r = 0, 1, 2, \dots, n-1 \right\}$ be the group of n th roots of unity. Let $a = e^{\frac{2\pi i}{n}}$, then $a \in G_n$. Since for any integer r , $e^{\frac{2\pi i r}{n}} = \left(e^{\frac{2\pi i}{n}} \right)^r = a^r$, every element of G_n is a power of a . Hence G_n is a cyclic group generated by a .

EXAMPLE 23. The additive group $\mathbf{Z}$ is a cyclic group generated by 1, since $1 \in \mathbf{Z}$ and for every integer n , we have $n = n \cdot 1$.

Theorem 2.33. A group of order n is cyclic if and only if it has an element of order n .

Proof: Let G be a cyclic group of order n and let a be a generator of G . Put $H = \{a^i \mid i \in \mathbf{Z}\}$. Clearly H is a subgroup of G . As G is finite a cannot be of infinite order. Let $o(a) = m$. We claim that H has m elements. Now $a, a^2, a^3, \dots, a^{m-1}, a^m = e$ all belong to H and no two of them are equal as $o(a) = m$. Hence H has at least m elements. Let $n \in H$ then $x = a^j, j \in \mathbf{Z}$. Now $j = mk + r, 0 \leq r < m, k$ and r integers. Then $x = a^j = (a^m)^k a^r = a^r$. So any element of H is one of $a, a^2, a^3, \dots, a^m$. Hence $H = \{a, a^2, a^3, \dots, a^m\}$. This proves our claim. Since $G = \langle a \rangle, G \subseteq H$. In other words $G = H$ and so $n = o(G) = o(H) = m = o(a)$.

Conversely let G be a group of order n and $b \in G$ be of order n . As before $K = \{b^r \mid r \in \mathbf{Z}\}$ is a subgroup of G having n elements as $o(b) = n$. Since $K \subseteq G$ and $o(K) = n = o(G)$, $K = G$. Consequently G is cyclic group generated by b . ■

Now let G be any group and $a \in G$. Let $H = \{a^n \mid n \text{ is any integer}\}$. Consider any two elements $x = a^n, y = a^m$ in H . We find that $xy^{-1} = a^n a^{-m} = a^{n-m} \in H$. This implies that H is subgroup of G . Clearly H is a cyclic group generated by a . This subgroup is called a *cyclic subgroup* of G generated by a and we write $H = \langle a \rangle$.

Theorem 2.34. If G is a finite group then order of any element of G divides the order of G .

Proof: Let G be a finite group and $a \in G$. Let $H = \langle a \rangle$ be a cyclic subgroup of G generated by a .

Then $o(H) = o(a)$ (Theorem 2.33).

By Lagrange's theorem $o(H) \mid o(G)$. Hence $o(a) \mid o(G)$. ■

Corollary 2.35. If G is a group finite group then for any $a \in G$
 $a^{o(G)} = e$.

Proof: Let $o(a) = n$ and $o(G) = N$.

Because of Theorem 2.34, $N = nm$ for some positive integer m . Then $a^N = (a^n)^m = e^m = e$.

Hence the corollary. ■

Theorem 2.36. In a group G , following hold:

- (1) For any two elements $a, x \in G$, $o(a) = o(x^{-1}ax)$.
- (2) If for any $a \in G$, $o(a)$ is finite; then for any integer m , $a^m = e$ implies $o(a) \mid m$.
- (3) For any two elements $a, b \in G$, $o(ab) = o(ba)$.
- (4) If $o(a) = n$ and a positive integer $k \mid n$, then

$$o(a^k) = \frac{n}{k}$$

Proof: (1) Let n be any positive integer

$$a^n = e \Leftrightarrow x^{-1} a^n x = x^{-1} e x = e \Leftrightarrow (x^{-1} a x)^n = e,$$

since $(x^{-1} a x)^n = x^{-1} a^n x$.

Consequently $o(a) = o(x^{-1} a x)$.

(2) Let $o(a) = n$ and let m be any integer such that $a^m = e$.

Now by Euclid's algorithm $m = nq + r$ for some integers q and r with $0 \leq r < n$.

$$\text{Then } e = a^m = (a^n)^q a^r = e^q a^r = e a^r = a^r.$$

Now n is the least positive integer such that $a^n = e$, so $r < n$ is absurd.

Hence $r = 0$ and so $m = nq$ i.e., $o(a) = n \mid m$. This proves (2).

(3) Since $ab = b^{-1}(ba)b$, $o(ab) = o(ba)$ by part (1).

(4) follows from (2). ■

EXAMPLE 24. Consider the group $G = \{1, -1, i, -i\}$ where $i = \sqrt{-1}$. Now $o(G) = 4$.

(i) $i^4 = 1$ but for no positive integer $m < 4$ $i^m = 1$.

This means $o(i) = 4$. Trivially $4 \mid o(G)$.

(ii) Now $(-1)^2 = 1$ and $(-1)^1 \neq 1$. This implies $o(-1) = 2$.

Clearly $o(1) \mid o(G)$.

(iii) Since $1 = i^4$, $-1 = i^2$, $-i = i^3$, $i = i$ we see that G is a cyclic group generated by i . Notice that G is also generated by $-i$. But -1 or 1 do not generate G .

Theorem 2.37. For any two subgroups H and K of a group G following hold:

(1) $H \cap K$ is a subgroup of G .

(2) If H is normal in G then $H \cap K$ is normal in K .

(3) If H and K are both normal in G , then $H \cap K$ is normal in G .

Proof: (1) Since $e \in H \cap K$, $H \cap K$ is non-void.

$$\begin{aligned} \text{Now } a, b \in H \cap K &\Rightarrow a, b \in H \text{ and } a, b \in K \\ &\Rightarrow ab^{-1} \in H \text{ and } ab^{-1} \in K \\ &\Rightarrow ab^{-1} \in H \cap K. \end{aligned}$$

Hence $H \cap K$ is a subgroup of G .

(2) Let H be normal in G . Let $x \in K$, $a \in H \cap K$.

Then $x^{-1} a x \in K$ since $x, a \in K$.

Further $x^{-1} a x \in H$ since H is normal and $a \in H$. Consequently $x^{-1} a x \in H \cap K \forall x \in K, a \in H \cap K$.

Hence $H \cap K$ is a normal subgroup of K .

(3) Let H and K be normal subgroups of G .

Now $x \in G$, $a \in H \cap K \Rightarrow x \in G, a \in H, a \in K$

$$\Rightarrow x^{-1} a x \in H \text{ and } x^{-1} a x \in K \text{ since } H \text{ and } K \text{ both are}$$

normal subgroups of G

This implies $x^{-1}ax \in H \cap K$.

Hence $H \cap K$ is a normal subgroups of G . ■

Definition 2.38. If A and B are two non-void subsets of a group $\langle G, * \rangle$ then by $A * B$ we mean the set $\{a * b \mid a \in A, b \in B\}$. If the composition in G is multiplication then $A * B$ is denoted by AB and is called the *product* of A and B . If the composition in G is addition, then $A * B$ is denoted by $A+B$ and is called the *sum* of A and B .

Theorem 2.39. Let H and K be two subgroups of a group G , then HK is a subgroup of G if and only if $HK=KH$.

Proof: Let HK be a subgroup of G . Consider $h \in H, k \in K$, $(kh)^{-1} = h^{-1}k^{-1} \in HK$; so $kh \in HK$ as HK is a subgroup of G . Consequently $KH \subseteq HK$.

Again $hk \in HK \Rightarrow (hk)^{-1} \in HK \Rightarrow (hk)^{-1} = h_1k_1$ for some $h_1 \in H, k_1 \in K \Rightarrow hk = (h_1k_1)^{-1} = k_1^{-1}h_1^{-1} \in KH$

Thus $HK \subseteq KH$. Hence $HK=KH$.

Conversely let $HK=KH$. Clearly $e \in HK$, so HK is non-empty. Now let $x, y \in HK$ then $x = h_1k_1$ and $y = h_2k_2$ for some $h_1, h_2 \in H, k_1, k_2 \in K$.

Thus $xy^{-1} = h_1k_1(k_2^{-1}h_2^{-1}) = h_1(k_1k_2^{-1})h_2^{-1}$.

Now $(k_1k_2^{-1})h_2^{-1} \in KH$ and by supposition $KH=HK$.

Consequently $(k_1k_2^{-1})h_2^{-1} = h_3k_3$ for some $h_3 \in H, k_3 \in K$.

We get $xy^{-1} = h_1(h_3k_3) = (h_1h_3)k_3 \in HK$.

This means that $x, y \in HK \Rightarrow xy^{-1} \in HK$.

Hence HK is a subgroup of G . ■

Corollary 2.40. If H and K are two subgroups of G such that either H or K is normal in G then HK is a subgroup of G .

Proof: Let H be a normal subgroup of G . Then for each $k \in K$, $Hk = kH$. This gives $HK = KH$.

Hence KH is a subgroup of G . Similarly if K is normal then for each $h \in H$, $hK = Kh$ which in turn implies $HK = KH$.

Hence the result follows from the above theorem. ■

Corollary 2.41. If H and K are two subgroups of an Abelian group G . Then HK is a subgroup of G . ■

REMARK 1. In Chapter 3, [§ 3, Problem 8], we give an example of two subgroups H and K of a group G for which $HK \neq KH$. Thus in general HK is not a subgroup.

REMARK 2. In Chapter 3, [§ 3, Problem 7, (iii)], we give an example of two subgroups H and K of a group G in which $HK = KH$ but neither of H and K is normal in G . (We advise the readers that they should not make an attempt to understand and solve the problems

referred above before they have grasped thoroughly the article on permutation groups.)

Theorem 2.42. *If H and K are finite subgroups of a group G then*

$$o(HK) = \frac{o(H) o(K)}{o(H \cap K)}.$$

Proof: Let $D = H \cap K$ and let $K = \bigcup_{i=1}^t Dk_i$ be a decomposition of K into disjoint right cosets of D in K . Then $t = \frac{o(K)}{o(D)}$. Further

$HK = H \left(\bigcup_{i=1}^t Dk_i \right) = \bigcup_{i=1}^t HK_i$ since $HD = H$ follows from $D \subseteq H$. Now we claim that the right cosets $Hk_1, Hk_2, \dots, Hk_t$ are pairwise distinct. Otherwise let $Hk_i = Hk_j$ for some i and j . This gives $k_i k_j^{-1} \in H$, but as $k_i k_j^{-1} \in K$, therefore $k_i k_j^{-1} \in H \cap K = D$. Consequently $Dk_i = Dk_j$ and $i = j$.

Hence $Hk_1, Hk_2, \dots, Hk_t$ are distinct and so disjoint right cosets of H . Each of them has $o(H)$ number of elements. We get

$$o(HK) = o(H)t = \frac{o(H) \cdot o(K)}{o(D)} = \frac{o(H)o(K)}{o(H \cap K)}. \blacksquare$$

WORKED-OUT EXERCISES

Exercise 1. *If H and K are two subgroups of a group G then for any $a, b \in G$ either $Ha \cap Kb = \phi$ or $Ha \cap Kb = (H \cap K)c$ for some $c \in G$.*

Solution. Suppose that $Ha \cap Kb$ is not empty. Let $c \in Ha \cap Kb$. But $c \in Hc$ and $c \in Kc$. Thus $c \in Ha \cap Hc$ and also $c \in Kb \cap Kc$. This in turn implies that $Ha = Hc$, $Kb = Kc$. Hence $Ha \cap Kb = Hc \cap Kc$.

Now $H \cap K \subseteq H \Rightarrow (H \cap K)c \subseteq Hc$

and $H \cap K \subseteq K \Rightarrow (H \cap K)c \subseteq Kc$.

Thus we get $(H \cap K)c \subseteq Hc \cap Kc$.

Again $t \in Hc \cap Kc \Rightarrow t = hc = kc$ for some $h \in H, k \in K$.

As a consequence $h = tc^{-1} = k \in H \cap K$.

So $t \in (H \cap K)c$.

This gives that $Hc \cap Kc \subseteq (H \cap K)c$.

Hence $Ha \cap Kb = Hc \cap Kc = (H \cap K)c$.

Exercise 2. (Poincaré) *If H and K are two subgroups of finite indices in G , then show that $H \cap K$ is also of finite index in G .*

Solution. Let $[G : H] = m$ and $[G : K] = n$ and let $Ha_1, Ha_2, \dots, Ha_m$ and $Kb_1, \dots, Kb_n$ be distinct right cosets of H and K res-

is a subgroup of G if and only if $HH \subseteq H$.

2. Give an example to show that a group G may have a non-void subset H for which $HH \subseteq H$ but H fails to be a subgroup of G .

3. Show that if H and K are normal subgroups of a group G then HK is a normal subgroup of G . Show further that the result does not necessarily hold if only one of them is normal.

4. For a group G and a subgroup H of G , let $G = Ha_1 \cup \dots \cup Ha_t$ be a decomposition of G into disjoint right cosets of H . Show that $G = a_1^{-1}H \cup \dots \cup a_t^{-1}H$ is a decomposition of G into left cosets of H .

5. Give an example of a group G having a subgroup H and two elements a, b such that $Ha = Hb$ but $aH \neq bH$.

6. If $H \subseteq K$ are two subgroups of a finite group G , then show that $[G : H] = [G : K][K : H]$

Hint. By Lagrange's Theorem, $o(G) = o(K)[G : K]$
 $o(K) = o(H)[K : H]$ and $o(G) = o(H)[G : H].$

7. Let $\mathbf{Z}$ be the additive group of integers and for any positive integer n , let H_n denote the set of all multiples of n . Show the following:

(i) H_n is a subgroup of $\mathbf{Z}$.

(ii) Index of H_n in $\mathbf{Z}$ is n .

(iii) For any two positive integers m, n , if j and k are their HCF and LCM respectively, then

$$H_j = H_n + H_m \text{ and } H_k = H_m \cap H_n.$$

8. Let H be a subgroup of a group G . Prove the following:

(i) For any $x \in G$, $x^{-1}Hx = \{x^{-1}hx \mid h \in H\}$ is a subgroup of G .

[Hint. Use Theorem 2.16.]

(ii) $o(H) = o(x^{-1}Hx)$.

[Hint. Show that $h \leftrightarrow x^{-1}hx$ is a 1-1 correspondence between elements of H and those of $x^{-1}Hx$.]

(iii) If G is finite and $n = o(H)$ such that G contains no subgroups other than H of order n , then prove that H is normal in G .

9. If $a \in G$ define $N(a) = \{x \in G \mid ax = xa\}$.

Prove the following:

(i) $N(a)$ is a subgroup of G .

(ii) For any two elements $x, y \in G$, $x^{-1}ax = y^{-1}ay$
 $\Leftrightarrow N(a)x = N(a)y$.

[Hint. Use the fact that for any subgroup H , $Hx = Hy \Leftrightarrow xy^{-1} \in H$.]

Definition: $N(a)$ is called the *Centralizer* or *Normalizer* of a .

10. If G is a group, then the *centre* $Z(G)$ of G is defined by $\{z \in G \mid zx = xz \ \forall \ x \in G\}$. Prove that Z is a normal subgroup

of G .

11. Show by an example that if H is a normal subgroup of G and K is a normal subgroup of H then K may not be a normal subgroup of G .

12. Give example of a group G and a subgroup H of G such that for some $a \in G$ $aHa^{-1} \subsetneq H$.

13. Let H and K be normal subgroups of G such that $H \cap K = \{e\}$. Show that $hk = kh \forall h \in H, k \in K$.

[Hint. Show that $hkh^{-1}k^{-1} \in H \cap K$.]

14. H is a non-empty subset of a group G such that $Hx = HxH$ for all $x \in G$, show that $Hx = xH$.

15. Let a, b be two elements of a group G such that $ab = ba$. Prove that if $o(a)$ and $o(b)$ are relatively prime integers then $o(ab) = o(a)o(b)$. Show by examples that the result is not true when (i) $ab \neq ba$ (ii) $ab = ba$ but $o(a)$ and $o(b)$ are not relatively prime integers.

16. If a cyclic subgroup T of a group G is normal in G , then show that every subgroup of T is normal in G .

17. Show that every cyclic subgroup of a group G is normal then every subgroup of G is normal.

18. If a is an element of G of finite order and k , any non-zero integer, then show that $o(a) = o(a^k)$ if and only if k and $o(a)$ are relatively prime.

19. Show that every group of prime order is cyclic.

20. Show that a subgroup H of group G is normal in G if and only if the product of any two right cosets of H in G is again a right coset of H in G .

21. If G has only one element a of order n , prove that $a \in Z(G)$.

22. Let H and K be two normal subgroups of a finite group G such that $o(H)$ and $o(K)$ are co-prime. Show that for any $h \in H, k \in K, hk = kh$.

[Hint. Use Lagrange's Theorem.]

23. (H.B. Mann) If G is a finite group and A, B are two non-void subsets of G then prove that either $G = AB$ or $o(G) \geq o(A) + o(B)$.

24. Show that if a group G has finite number of subgroups, then G is finite.

25. Let A and B be subgroups of group G . If $[G : B]$ is finite then prove that $[A : A \cap B] \leq [G : B]$. Equality holds if and only if $G = AB$.

26. In problem 25 above, prove that if $[G : A]$ is also finite then

$[G : A \cap B] \leq [G : A][G : B]$ with equality if and only if $G = AB$.

27. If in a group G , $a^3 = e$ and $aba^{-1} = b^2$ for $a, b \in G$, prove that if $b \neq e$ then $o(b) = 31$.

28. If H and K are subgroups of a finite group G and $o(H) > \sqrt{o(G)}$ and $o(K) > \sqrt{o(G)}$, show that $o(H \cap K) > 1$.

29. (Euler) If n is a positive integer and a is relatively prime to n , then show that $a^{\phi(n)} \equiv 1 \pmod{n}$ where $\phi(n)$ is the number of positive integers less than n and relatively prime to n .

30. (Fermat) If p is a prime integer and a is any integer, then prove that $a^p \equiv a \pmod{p}$.

31. Let $\mathbf{Q}$ be the set of all rational numbers and let $G = \{(a, b) \mid a, b \in \mathbf{Q}, a \neq 0\}$. Define $*$ on G as follows:

$$(a, b) * (c, d) = (ac, ad + b).$$

Show that $\langle G, * \rangle$ is a non-Abelian group.

Consider $H = \{(1, b) \mid b \in \mathbf{Q}\}$. Prove that H is a normal subgroup of G .

32. If a group G is the union of a family of proper normal subgroups each two of which have only identity in common, then show that G is Abelian.

33. If A and B are subgroups of finite index in a group G , and $[G : A]$ and $[G : B]$ are relatively prime, then show that $G = AB$.

[Hint. Use Problem 26 above.]

34. For a non-empty set S of a group G , define $S^k = \{\prod_{i=1}^k s_i \mid s_i \in S\}$ for any positive integer k . Prove that, if G has n elements, then S^n is a subgroup of G .

35. Show that if a group G is of even order then number of elements of order 2 is odd.

36. G is an Abelian group having n elements $g_1, g_2, \dots, g_n$. Show that $(g_1, g_2, \dots, g_n)^2 = e$, where e is identity of G .

37. Give example of a group G and elements $a, b \in G$ such that $o(a)$ and $o(b)$ are finite but ab is not of finite order.

[Hint: Consider the group $S(\mathbf{R})$, the set of all 1-1 onto functions from $\mathbf{R}$ to $\mathbf{R}$.

Define $a(x) = -x$ and $b(x) = 1 - x$ for all $x \in \mathbf{R}$. Clearly $o(a) = o(b) = 2$, but ab is not of finite order.]

QUOTIENT GROUPS, HOMOMORPHISMS AND PERMUTATIONS

Introduction

We have seen in the last chapter that there are large number of examples of groups of various types. In this chapter we attempt to do the following:

(i) We introduce the notion of homomorphism so that we may establish relationship between various groups; (ii) give some methods of constructing new groups from known groups; (iii) study groups of permutations, and in the last i.e (iv) we introduce the conjugacy among elements of a group so that we may explore the inner structure of a group.

1. Quotient Groups

Let G be a group and N , a normal subgroup of G . We know that $Nx = xN \quad \forall \quad x \in G$ i.e. the right and left cosets of N determined by any $x \in G$ are equal. So there is no need to distinguish between left and right cosets of a normal subgroup N . We shall say that Nx is a coset of N . Let G/N denote the set of all cosets of N in G .

Theorem 3.1. Let G be a group and N be a normal subgroup of G . The set G/N of all cosets of N in G , is a group under the binary composition defined as under. For any $Na, Nb \in G/N$,

$$(Na)(Nb) = Nab.$$

Proof: Before we go into the details of the proof, we make a remark about the composition defined above. Apparently it looks that the product $(Na)(Nb) = Nab$, depends on the choice of the representatives a and b of Na and Nb respectively. However if we define the product of two cosets, we will like that this product should not depend upon a and b , but should depend on the cosets Na and Nb themselves. It can happen that one may find four elements, a, a_1, b, b_1 such that $a \neq a_1, b \neq b_1$, but $Na = Na_1, Nb = Nb_1$. By the

formula for product $(Na_1)(Nb_1)=Na_1b_1$ and $(Na)(Nb)=Nab$. As $Na=Na_1$ and $Nb=Nb_1$, we must naturally then expect that $(Na)(Nb)=(Na_1)(Nb_1)$ i.e., $Nab=Na_1b_1$; otherwise the composition defined above will have no sense.

Now we come to the actual proof.

Let $Na=Na_1$, $Nb=Nb_1$.

This $\Rightarrow a_1=n_1a$, $b_1=n_1b$ for some $n_1, n_2 \in N$,

$$\Rightarrow a_1b_1=(n_1a)(n_2b)=n_1a(n_2a^{-1})(ab).$$

As N is normal, $an_2a^{-1} \in N$. This yields $n_1an_2a^{-1} \in N$. Consequently $N(n_1an_2a^{-1})=N$.

Then $Na_1b_1=N(n_1an_2a^{-1})ab=Nab$. This proves that the binary composition is well-defined.

Let $Nu, Nb, Nc \in G/N$.

(1) *Associativity*. Now $[(Na)(Nb)](Nc)=(Nab)(Nc)=N(ab)c$.

Also $(Na)[(Nb)(Nc)]=Na(bc)$.

But $a(bc)=(ab)c$. Hence $[(Na)(Nb)](Nc)=(Na)[(Nb)(Nc)]$.

(2) *Existence of Identity*. Let e be the identity of G .

Then $N=Ne \in G/N$.

Now $(Na)N=(Na)(Ne)=Nae=Na$. Similarly $(N)(Na)=Na$.

This shows that N is the identity of G/N .

(3) *Existence of Inverse*.

Since $(Na)(Na^{-1})=Naa^{-1}=Ne=Na^{-1}a=(Na^{-1})(Na)$,

we get $(Na)^{-1}=Na^{-1} \in G/N$.

Hence G/N is a group. ■

Definition 3.2. Let G be a group and let N be a normal subgroup of G . Then the set G/N of all cosets of N in G , together with the binary composition defined by $(Na)(Nb)=Nab$ for all $Na, Nb \in G/N$, is a group, and is called the *quotient group* of G by N .

A quotient group is also called a *factor group*. Further if the binary composition in G is addition, each coset of N in G is denoted as $N+a$, then the binary composition in G/N is also denoted additively i.e. we write $(N+a)+(N+b)=N+(a+b)$.

Lemma 3.3. For a finite group G and for each normal subgroup N of G , $o(G/N)=o(G)/o(N)$.

Proof: Let t be the number of distinct cosets of N in G . Then by Lagrange's Theorem $o(G)=o(N)t$.

Hence $t=\frac{o(G)}{o(N)}$ i.e., $o(G/N)=o(G)/o(N)$. ■

EXAMPLE 1. Let $\mathbf{Z}$ be the additive group of integers. Consider $N=\langle 3 \rangle$, the subgroup of $\mathbf{G}$ consisting of all multiples of 3. Now $N, N+1, N+2$ are three cosets of N in $\mathbf{Z}$. Firstly we show that these are the only three cosets of N . For, let a be any integer, by division

algorithm $a=3b+c$ for some $b, c \in \mathbb{Z}$ such that $0 \leq c < 3$. Then $a-c=3b \in N$ gives $N+a=N+c$. As $c=0, 1$ or 2 , $N+a$ is one of $N, N+1, N+2$. Further these three cosets are distinct, as for no two distinct non-negative integers a and b both less than 3 , $a-b$ is a multiple of 3 i.e. $N+a=N+b$. Hence $\mathbb{Z}/N=\{N, N+1, N+2\}$. As $(N+1)+(N+2)=N+3=N$ (since $3 \in N$), $-(N+1)=N+2$.

PROBLEMS

1. (a) Let G be an Abelian group and N , a subgroup of G . Prove that the quotient group G/N is Abelian.

(b) Give an example of a non-Abelian group G and a proper normal subgroup N of G , such that G/N is Abelian. [Hint. Consider Quaternion group, problem 9 at the end of Section 2.2.]

2. (a) Using Lagrange's Theorem, prove that if G is a finite group, and N , a normal subgroup of G , then for any $x \in G$, $o(x)$ in G is divisible by the order of $\bar{x}=Nx$ in G/N .

(b) If N is a normal subgroup of a finite group G such that $o(N)$ and $[G : N]$ are co-prime, show that for any $x \in G$, $x^{o(N)}=e \Rightarrow x \in N$.

3. Let $G=\{1, -1, i, -i\}$ and $N=\{1, -1\}$ where $i=\sqrt{-1}$. Show that N is normal subgroup of G and construct the multiplication table of the quotient group G/N .

4. Let n be a positive integer and $N=\langle n \rangle$ be the cyclic subgroup of the additive group $\mathbb{Z}$ of integers. Show that $o(\mathbb{Z}/N)=n$.

2. Homomorphisms

Definition 3.3. (Homomorphism). Let G and G' be any two groups and let ' $\circ$ ' and ' $*$ ' denote their respective binary compositions. Then a mapping $f: G \rightarrow G'$ is called a *homomorphism* if

$$f(a \circ b) = f(a) * f(b) \quad \forall a, b \in G.$$

A homomorphism f , which at the same time is also onto is called an *epimorphism*. A homomorphism f which is at the same time 1-1 is called a *monomorphism*.

A group G' is said to be a homomorphic image of a group G , if there exists a homomorphism f of G onto G' . A homomorphism of a group G into itself is called an *endomorphism* of G .

EXAMPLE 2. Consider $\mathbb{Z}$, the additive group of integers and $G=\{2^n : n \in \mathbb{Z}\}$.

G is a group under the usual multiplication of real numbers.

Define $f: \mathbb{Z} \rightarrow G$ by $f(n)=2^n \quad \forall n \in \mathbb{Z}$.

Since $f(n+m)=2^{n+m}=2^n \cdot 2^m \quad \forall n, m \in \mathbb{Z}$, f is a homomorphism of $\mathbb{Z}$ into G . Clearly f is onto and G is a homomorphic image of $\mathbb{Z}$.

EXAMPLE 3. Let G be the multiplicative group of all non-singular $n \times n$ matrices over the real numbers. Let R^* be the multiplicative group of all non-zero real numbers.

Define $f: G \rightarrow R^*$ by $f(A) = \det A$ for all $A \in G$.

Since for any two $n \times n$ matrices A, B , $\det(AB) = \det A \cdot \det B$, we get $f(AB) = f(A)f(B)$.

Hence f is a homomorphism of G into R^* . f is also onto (verify!).

EXAMPLE 4. Let $G = \{1, -1\}$. G is a group under multiplication. Define $f: \mathbb{Z} \rightarrow G$ by putting $f(n) = 1$ if n is even, and $f(n) = -1$ if n is odd. Consider any two integers m, n . Following cases arise.

Case I. Both m, n are even. Then $m+n$ is even. Consequently by definition $f(m) = 1, f(n) = 1, f(m+n) = 1$.

This yields $f(m+n) = f(m)f(n)$.

Case II. One of m and n is even and other is odd. To be definite let us suppose m is even and n is odd.

Then $f(m) = 1$ and $f(n) = -1$.

Further as $m+n$ is also odd, $f(m+n) = -1 = 1(-1) = f(m)f(n)$.

Case III. Both of m and n are odd.

Then $f(m) = -1$ and $f(n) = -1$.

Now as $m+n$ is even, $f(m+n) = 1 = (-1)(-1) = f(m)f(n)$.

This shows that $f(m+n) = f(m)f(n) \forall m, n \in \mathbb{Z}$.

Since $f(2) = 1, f(3) = -1$,

we see that f is also onto.

Hence f is an epimorphism; consequently G is a homomorphic image of $\mathbb{Z}$.

EXAMPLE 5. Let E be the additive group of even integers. Then the mapping $f: \mathbb{Z} \rightarrow E$ in which $f(n) = 2n \forall n \in \mathbb{Z}$ is a homomorphism. Since each member of E is of the form $2n, n \in \mathbb{Z}$, we get that f is onto. Hence E is a homomorphic image of $\mathbb{Z}$. f is also 1-1 mapping (?).

Following our usual convention of denoting the binary composition in a group by multiplication, we can say that a mapping f of a group G into a group G' is a homomorphism if

$$f(ab) = f(a)f(b) \forall a, b \in G.$$

Definition 3.4. (Kernel of a homomorphism).

Let f be a homomorphism of a group G into a group G' , then the kernel of f is the set $= \{x \in G \mid f(x) = e', \text{ the identity of } G'\}$.

We shall denote the kernel of f by $\text{Ker } f$.

Theorem 3.5. Let G and G' be any two groups, e and e' , their respective identities. If f is a homomorphism of G into G' , then

$$(1) f(e) = e'.$$

$$(2) f(x^{-1}) = f(x)^{-1} \quad \forall x \in G.$$

(3) *Ker f* is normal subgroup of *G*.

Proof: (1) Since $e.e=e$, $f(e)f(e)=f(e)$. However $f(e) \in G'$ gives $f(e)=e'f(e)$. Thus $f(e)f(e)=e'f(e) \Rightarrow f(e)=e'$, by right cancellation in G' .

(2) For any $x \in G$, since $xx^{-1}=e$, we get $f(x)f(x^{-1})=f(xx^{-1})=f(e)=e'$.

Similarly $x^{-1}x=e$ gives $f(x^{-1})f(x)=e'$.

Hence $f(x^{-1})=f(x)^{-1}$.

(3) Since $f(e)=e'$, $e \in \text{Ker } f$.

This shows $\text{Ker } f \neq \phi$.

Now let $a, b \in \text{Ker } f$, $x \in G$. $a \in \text{Ker } f$, $b \in \text{Ker } f \Rightarrow f(a)=e'$, $f(b)=e'$

$$\Rightarrow f(a)=e', f(b^{-1})=f(b)^{-1}=e'$$

$$\Rightarrow f(ab^{-1})=f(a)f(b)^{-1}=e'e'=e'$$

$$\Rightarrow ab^{-1} \in \text{Ker } f.$$

This proves that $\text{Ker } f$ is a subgroup of G . It remains to show that it is also normal.

$$\begin{aligned} \text{Now } f(x^{-1}ax) &= f(x^{-1})f(a)f(x) \\ &= f(x)^{-1}f(a)f(x) \\ &= f(x)^{-1}e'f(x) \\ &= f(x)^{-1}f(x)=e'. \end{aligned}$$

Consequently $x^{-1}ax \in \text{Ker } f$.

Hence $\text{Ker } f$ is a normal subgroup of G . ■

Definition 3.6. (Isomorphism). A homomorphism f of a group G onto a group G' is called an *isomorphism* if f is a 1-1 mapping.

It is now clear that an isomorphism is a homomorphism which is simultaneously an epimorphism and a monomorphism.

Further an isomorphism of a group G onto itself is called an *automorphism* of G .

EXAMPLE 6. For any group G , the identity mapping $i: G \rightarrow G$ is an automorphism of G .

Since by definition $i(x)=x \forall x \in G$, i is clearly 1-1 as well as onto. Further for any $x, y \in G$, $i(xy)=xy=i(x)i(y)$ gives i is an automorphism of G .

EXAMPLE 7. Let R^+ be the multiplicative group of all strictly positive real numbers. Define $f: R^+ \rightarrow R^+$ by $f(x)=x^2 \forall x \in R^+$.

Now for any $x, y \in R^+$, $f(xy)=(xy)^2=x^2y^2=f(x)f(y)$.

Thus f is an endomorphism of R^+ .

Further $f(x)=f(y) \Rightarrow x^2=y^2 \Rightarrow x=y$, since $x > 0, y > 0$.

Hence f is a 1-1 mapping.

Given $x \in R^+$, $\sqrt{x} \in R^+$ such that $f(\sqrt{x})=(\sqrt{x})^2=x$.

This proves that f is also onto. Consequently f is an automorphism

of R^+ .

EXAMPLE 8. For a fixed $a \in G$, define

$$f_a : G \rightarrow G \text{ by } f_a(x) = axa^{-1} \quad \forall x \in G.$$

For any $x, y \in G$, $f_a(xy) = axya^{-1}$

$$= (axa^{-1})(aya^{-1}) = f_a(x)f_a(y).$$

$$f_a(x) = f_a(y) \Rightarrow axa^{-1} = aya^{-1} \Rightarrow x = y.$$

Since $f_a(a^{-1}xa) = (a^{-1}xa)a^{-1} = x$, we get, f is also onto. Hence f is an automorphism of G .

Further the homomorphisms in Examples 2, 5 are isomorphisms while those of Examples 3, 4 are not isomorphisms.

Theorem 3.7. A homomorphism of a group G into a group G' is a monomorphism if and only if $\text{Ker } f = \{e\}$.

Proof: Let f be a homomorphism of a group G into a group G' . Suppose that f is 1-1. Let $x \in \text{Ker } f$. Then $f(x) = e'$, the identity of G' . Also $f(e) = e'$, (Theorem 3.5). Thus $f(x) = f(e) \Rightarrow x = e$. This proves that $\text{Ker } f = \{e\}$.

Conversely let $\text{Ker } f = \{e\}$.

Let $x, y \in G$, be such that $f(x) = f(y)$.

$$\text{Then } f(x)f(y)^{-1} = e' \Rightarrow f(xy^{-1}) = f(x)f(y^{-1}) = e'$$

$$\Rightarrow xy^{-1} \in \text{Ker } f = \{e\}$$

$$\Rightarrow xy^{-1} = e \Rightarrow x = y.$$

Hence f is 1-1. ■

Definition 3.8. A group G is said to be isomorphic to a group G' if there exists an isomorphism of G onto G' .

Symbolically $G \cong G'$ will denote that G is isomorphic to G' .

Theorem 3.9. Let G_1, G_2, G_3 be any three groups.

Then

$$(1) G_1 \cong G_1 \text{ (Reflexivity)}$$

$$(2) G_1 \cong G_2 \Rightarrow G_2 \cong G_1 \text{ (Symmetry)}$$

$$(3) G_1 \cong G_2, G_2 \cong G_3 \Rightarrow G_1 \cong G_3 \text{ (Transitivity)}.$$

Proof: (1) Define $i : G_1 \rightarrow G_1$ such that $i(x) = x \quad \forall x \in G_1$. Then i , the identity map on G_1 , is an automorphism of G_1 (Example 6).

Hence $G_1 \cong G_1$.

(2) $G_1 \cong G_2 \Rightarrow \exists$ an isomorphism f of G_1 onto G_2 . Since f is 1-1 as well as onto, the mapping $f^{-1} : G_2 \rightarrow G_1$ given by $f^{-1}(y) = x$ if $f(x) = y$ for any $y \in G_2$, is 1-1 as well as onto. Let $y_1, y_2 \in G_2$. Since f is onto, we can find $x_1, x_2 \in G_1$ such that $f(x_1) = y_1$ and $f(x_2) = y_2$.

$$\text{This } \Rightarrow f^{-1}(y_1) = x_1, f^{-1}(y_2) = x_2.$$

$$\text{However } f(x_1x_2) = f(x_1)f(x_2) = y_1y_2 \text{ gives } f^{-1}(y_1y_2) = x_1x_2 = f^{-1}(y_1)f^{-1}(y_2).$$

This proves that f^{-1} is also a homomorphism. Hence f^{-1} is an iso-

morphism of G_2 onto G_1 and $G_2 \cong G_1$.

(3) Let $G_1 \cong G_2$, $G_2 \cong G_3 \Rightarrow \exists$ an isomorphism f of G_1 onto G_2 and an isomorphism g of G_2 onto G_3 . Since f and g are both 1-1 and onto mappings, $gof : G_1 \rightarrow G_3$ is also a 1-1 as well as onto mapping. Now for any $x, y \in G_1$, $(gof)(xy) = g(f(xy))$

$$= g(f(x)f(y))$$

$$= g(f(x))g(f(y)), \text{ since } f(x), f(y) \in G_2,$$

$$= [(gof)(x)][(gof)(y)].$$

This proves that gof is a homomorphism. Hence gof is an isomorphism of G_1 onto G_3 and $G_1 \cong G_3$.

This proves the theorem. ■

REMARKS. (1) In the proof of part (3) of the above theorem, when we prove that gof is a homomorphism, we do not need in the proof that g and f are 1-1 or onto. Thus it follows that the composite (whenever defined) of two homomorphisms is again a homomorphism.

(2) The above theorem asserts that given any family of groups, the relation of isomorphism is an equivalence relation on the family. Thus if a group G_1 is isomorphic to a group G_2 then we simply say that G_1 and G_2 are isomorphic. When two groups are isomorphic then for an algebraist they are essentially equal because of the following reasons:

The algebraist is interested only in the study of binary composition on a group and not in the intrinsic nature of the elements of the group. So if $G_1 \cong G_2$ and f is the underlying isomorphism of G_1 onto G_2 then for any $x, y \in G$, $xy = z \Rightarrow f(x)f(y) = f(z)$. That means if f is known, x and y are known, then $f(x)f(y)$ being equal to $f(xy)$ is known in G_2 , the moment xy is known in G_1 . Consequently the binary composition in G_2 is known completely if the binary composition in G_1 and the mapping f are known. Similarly if we know the binary composition on G_2 and mapping g of G_2 onto G_1 which is an isomorphism, then the binary composition on G_1 is known completely. Thus if all the properties of one of the groups G_1 and G_2 , which are dependent only on its binary composition, are known, then all the properties of the other which are dependent upon its binary composition are also known. Because of these reasons, an algebraist regards two isomorphic groups to be identical.

Theorem 3.10. (Fundamental Theorem of Homomorphism). Let G be a group. If N is a normal subgroup of G , then G/N is a homomorphic image of G . Conversely if any group G' is a homomorphic image of G then G' is isomorphic to some quotient group of G . In fact, if f is a homomorphism of G onto G' , then $G' \cong G/\text{Ker } f$.

Proof: Define $f: G \rightarrow G/N$ by $f(a) = Na \quad \forall a \in G$.

Since $f(ab) = Nab = (Na)(Nb) = f(a)f(b) \quad \forall a, b \in G$, and for each $Nc \in G/N$, $f(c) = Nc$, we see that f is a homomorphism of G onto G/N .

Conversely let G' be a homomorphic image of G , then there exists a homomorphism g of G onto G' . Let $N = \text{Ker } g$. We know that N is a normal subgroup of G (Theorem 3.5). Consider the quotient group G/N . Define $f: G/N \rightarrow G'$ by $f(Na) = g(a) \quad \forall a \in G$.

Firstly we show that f is well-defined. For this, let $a_1, a_2 \in G$ with $Na_1 = Na_2$.

This $\Rightarrow a_1 a_2^{-1} \in N \Rightarrow g(a_1 a_2^{-1}) = e'$, the identity of G'
 $\Rightarrow g(a_1) g(a_2)^{-1} = e' \Rightarrow g(a_1) = g(a_2)$
 $\Rightarrow f(Na_1) = f(Na_2)$.

Consequently f is well-defined.

Now for any $Na, Nb \in G/N$

$$f[(Na)(Nb)] = f(Nab) = g(ab) = g(a)g(b) = f(Na)f(Nb).$$

This shows that f is a homomorphism.

Further

$$\begin{aligned} f(Na) = f(Nb) &\Rightarrow g(a) = g(b), \\ \Rightarrow g(a)g(b)^{-1} &= e' \Rightarrow g(ab^{-1}) = e' \\ \Rightarrow ab^{-1} &\in \text{Ker } g = N \Rightarrow Na = Nb. \end{aligned}$$

Thus f is a 1-1 mapping.

Lastly let $x \in G'$, as g is onto $\exists a \in G$ such that $g(a) = x$. Consequently $f(Na) = g(a) = x$. This proves that f is also onto. Hence f is an isomorphism of G/N onto G' and $G/N \cong G'$. This proves the theorem. ■

Definition 3.11. (Natural Homomorphism). Let N be a normal subgroup of a group G , then the mapping $f: G \rightarrow G/N$ defined by $f(x) = Nx \quad \forall x \in G$ is called the *natural or canonical homomorphism* of G onto G/N .

Theorem 3.12. (First Theorem of Isomorphism). Let f be a homomorphism of a group G onto a group G' and $H = \text{Ker } f$, K' any normal subgroup of G' and $K = \{x \in G \mid f(x) \in K'\} = f^{-1}(K')$. Then K is a normal subgroup of G containing H and $G/K \cong G'/K'$.

Proof: Define $g: G \rightarrow G'/K'$ by $g(x) = K'f(x) \quad \forall x \in G$.

Let $a, b \in G$. Then $g(ab) = K'f(ab) = K'f(a)f(b)$, (since f is a homomorphism), so $g(ab) = [K'f(a)][K'f(b)] = g(a)g(b)$. Consequently g is a homomorphism. Further let $K'a' \in G'/K'$, $a' \in G'$. As f is onto there exists $a \in G$ such that $f(a) = a'$. Then $g(a) = K'f(a) = K'a'$. So g is a homomorphism of G onto G'/K' . By the Fundamental Theorem of Homomorphism $G/\text{Ker } g \cong G'/K'$.

If we show that $\text{Ker } g = K$, then it will follow that K is a normal subgroup of G and $G/K \cong G'/K'$.

Now as K' is the identity of G'/K' ,

$$\begin{aligned} x \in \text{Ker } g &\Leftrightarrow g(x) = K' \\ &\Leftrightarrow K' f(x) = K' \\ &\Leftrightarrow f(x) \in K \\ &\Leftrightarrow x \in K, \text{ by definition of } K. \end{aligned}$$

This gives that $x \in \text{Ker } g \Leftrightarrow x \in K$; hence $\text{Ker } g = K$. Now $x \in H \Rightarrow f(x) = e' \in K' \Rightarrow g(x) = K' f(x) = K' \Rightarrow x \in \text{Ker } g = K \Rightarrow H \subseteq K$. This proves the theorem. ■

Corollary 3.13. (Freshman's theorem). Let H and K be two normal subgroups of G such that $H \subseteq K$. Then K/H is a normal subgroup of G/H and

$$G/K \cong \frac{G/H}{K/H}.$$

Proof: We leave it to the readers to verify that K/H is a normal subgroup of G/H . Further $f: G \rightarrow G/H$ defined by $f(x) = Hx$ is the natural homomorphism of G onto G/H such that $K = f^{-1}(K/H) = \{x \in G \mid Hx \in K/H\}$. By the above theorem

$$G/K \cong \frac{G/H}{K/H}. \quad \blacksquare$$

Theorem 3.14 (a) (correspondence Theorem). **READ** Let N be a normal subgroup of a group G .

$$K/(H \cap K) \cong HK/H.$$

Proof: Since H is a normal subgroup of G , $HK = KH$. Consequently HK is a subgroup of G . Further $H = He \subseteq HK$ gives that H is a normal subgroup of HK ; consequently HK/H is defined.

Define $f: K \rightarrow HK/H$ by $f(k) = Hk$. f is a homomorphism. Consider any $Ha \in HK/H$, $a \in HK$. Now $a \in HK$ gives $a = hk$ for some $h \in H$, $k \in K$.

Thus $Ha = (Hh)k = Hk = f(k)$. This shows that f is also onto. By the Fundamental Theorem of Homomorphism $K/\text{Ker } f \cong HK/H$.

If we can establish that $\text{Ker } f = H \cap K$, we shall get that $H \cap K$ is a normal subgroup of K and $K/(H \cap K) \cong HK/H$. Now H is the identity of HK . So for any $k \in K$, $k \in \text{Ker } f \Leftrightarrow f(k) = H$

$$\Leftrightarrow Hk = H \Leftrightarrow k \in H \Leftrightarrow k \in H \cap K, \text{ since } k \in K.$$

i.e. $k \in \text{Ker } f \Leftrightarrow k \in H \cap K$. This yields $\text{Ker } f = H \cap K$.

Hence that result follows. ■

Theorem 3.14 (a). (Correspondence Theorem). Let N be a subgroup of a group G . There is 1-1 correspondence between subgroups of G containing N and subgroups of G/N . This correspondence also gives a 1-1 correspondence between normal subgroups of G containing N and subgroups of G/N .

Proof: Let f be the natural homomorphism from G onto G/N . Let H be a subgroup of G containing N . Then $f(H) = \{f(h) \mid h \in H\} = \{hN \mid h \in H\} = H/N$ is a subgroup of G/N . (Why?) Consider any two subgroups H and K of G containing N such that $f(H) = f(K)$ i.e. $H/N = K/N$. Then for any $h \in H$, there exists $k \in K$ such that $hN = kN$. This implies that $k^{-1}h \in N \subseteq K \Rightarrow h = k(k^{-1}h) \in K$. Consequently $H \subseteq K$. Similarly it can be proved that $K \subseteq H$. Hence $H = K$. Thus we see that the mapping $H \rightarrow f(H)$ is 1-1. Let X be any subgroup of G/N , put $Y = \{x \in G \mid f(x) \in X\}$. Since for each $n \in N$, $f(n) = N \in X$ (N is identity of G/N), $n \in Y$. Thus $N \subseteq Y$. Also if $t \in X \subseteq G/N$, $t = gN = f(g)$, so $g \in Y$. In this way we see that each element of X is of type $f(y)$ with $y \in Y$. Now we show that Y is a subgroup of G . Clearly Y is non-empty (Recall that $N \subseteq Y$). If $y_1, y_2 \in Y$ then $f(y_1) \in X, f(y_2) \in X \Rightarrow f(y_1)[f(y_2)]^{-1} \in X \Rightarrow f(y_1)f(y_2^{-1}) \in X \Rightarrow f(y_1y_2^{-1}) \in X \Rightarrow y_1y_2^{-1} \in Y$. Thus Y is subgroup of G containing N . Since $f(Y) = X$, we get that the mapping $H \rightarrow f(H)$ is a 1-1 correspondence.

Finally let H be a normal subgroup of G containing N . Consider $f(H)$. Let $gN \in G/N, hN \in f(H)$ then

$$gN hN (gN)^{-1} = (ghg^{-1})N \in f(H) \text{ as } ghg^{-1} \in H.$$

In this way $f(H)$ is a normal subgroup of G/N .

Let K be a normal subgroup of G/N . Put $L = \{x \in G \mid f(x) \in K\}$. Then it can be seen as above that $N \subseteq L, K = f(L)$ and L is a subgroup of G . Let $g \in G, x \in L$ then $f(gxg^{-1}) = f(g)f(x)f(g)^{-1} \in K$ as $f(g) \in G/N, f(x) \in K$ and K is a normal subgroup of G/N . As a consequence $gxg^{-1} \in L$. This proves that L is a normal subgroup of G containing N . ■

WORKED-OUT EXERCISES

Exercise 1. For $a, b \in \mathbf{R}, a \neq 0$, define $f : \mathbf{R} \rightarrow \mathbf{R}$ by $f_{ab}(x) = ax + b$. Let $G = \{f_{ab} \mid a, b \in \mathbf{R}, a \neq 0\}$ and $N = \{f_{1b} \in G\}$. Prove that N is a normal subgroup of G and $G/N \cong$ the group of non-zero real numbers under multiplication.

Solution. $f_{10}(x) = x$ shows that the identity mapping $i_{\mathbf{R}} \in G$. So G is non-empty.

Let $a, b, c, d \in \mathbf{R}$ with $a \neq 0$ and $c \neq 0$

$$\text{then } f_{ab}[f_{cd}(x)] = f_{ab}(cx + d) = a(cx + d) + b = acx + ad + b$$

$$\text{Thus } f_{ab}f_{cd} = f_{ac, ad+b} \dots (1)$$

Also $f_{ac, ad+b} \in G$ as $a \neq 0, c \neq 0 \Rightarrow ac \neq 0$.

With the help of (1) it can be proved that G is a group with f_{10} working as its identity and $(f_{ab})^{-1} = f_{a^{-1}, -a^{-1}b}$.

Clearly $f_{10} \in N$, so N is non-empty. Further if

$$f_{1b}, f_{1d} \in N \text{ then } f_{1b} (f_{1d})^{-1} = f_{1b} f_{1d}^{-1} = f_{1, -d+b} = f_{1, -d+b} \in N$$

So N is a subgroup of G . Again if $f_{ab} \in G$ and $f_{1c} \in N$ then

$$f_{ab} f_{1c} (f_{ab})^{-1} = f_{ab} f_{1c} f_{ab}^{-1} = f_{1, ac+b} = f_{1, ac} \in N$$

Hence N is a normal subgroup of G .

Finally let K be the group of non-zero real numbers under multiplication, define $\theta : G \rightarrow K$ by $\theta(f_{ab}) = a$ for all $f_{ab} \in G$. Obviously θ is onto. Now

$$\theta(f_{ab} f_{cd}) = \theta(f_{ac, ad+b}) = ac = \theta(f_{ab}) \theta(f_{cd})$$

implies that θ is a homomorphism.

$$f_{ab} \in \text{Ker } \theta \Leftrightarrow \theta(f_{ab}) = \text{identity of } K$$

$$\Leftrightarrow a = 1$$

$$\Leftrightarrow f_{ab} = f_{1b}$$

$$\Leftrightarrow f_{ab} \in N$$

Hence $\text{Ker } \theta = N$.

By Fundamental Theorem of Homomorphism

$$\frac{G}{N} \cong K.$$

Exercise 2. If G is a group such that $G/Z(G)$ is cyclic, where $Z(G)$ is the centre of G . Show that G is Abelian.

Solution. Put $N = Z(G)$ and let $G/N = \langle gN \rangle$ for some $g \in G$.

Let $a, b \in G$ then $aN = (gN)^k = g^k N$

and $bN = (gN)^l = g^l N$ for some $k, l \in \mathbb{Z}$.

Thus $a = g^k n_1$ and $b = g^l n_2$, $n_1, n_2 \in N$.

This gives that $ab = g^{k+l} n_1 n_2$ and $ba = g^{l+k} n_1 n_2$ as $n_1, n_2 \in N = Z(G)$.

Hence $ab = ba$.

Exercise 3. If R^* is the group of non-zero real numbers under multiplication then show that $(R^*, \cdot)$ is not isomorphic to $(R, +)$.

Solution. $-1 \in R^*$ is of order 2. So if at all $(R^*, \cdot) \cong (R, +)$ say, under an isomorphism f ; then $f(-1)$ must be of order 2 (why?). But for every $a \in R$, $na = 0 \Rightarrow n = 0$ or $a = 0$.

So there is no element of order 2 in $(R, +)$. Hence $(R^*, \cdot)$ cannot be isomorphic to $(R, +)$.

Exercise 4. N is a normal subgroup of a group G . Show that G/N is abelian if and only if for all $x, y \in G$, $xyx^{-1}y^{-1} \in N$.

Solution. In case $xyx^{-1}y^{-1} \in N$ for all $x, y \in G$,

$$xyx^{-1}y^{-1}N = N \Rightarrow xNyN(xN)^{-1}(yN)^{-1} = N$$

$$\Rightarrow xNyN = yNxN \Rightarrow G/N \text{ is abelian.}$$

Conversely let G/N be abelian, then for any $x, y \in G$

$$xNyN(xN)^{-1}(yN)^{-1} = (xN)(xN)^{-1}(yN)(yN)^{-1} = NN = N.$$

$$\text{Now } xyx^{-1}y^{-1}N = xNyNx^{-1}Ny^{-1}N = xNyN(xN)^{-1}(yN)^{-1}N = N \Rightarrow xyx^{-1}y^{-1} \in N.$$

Exercise 5. If H is a subgroup of group G show that $W = \text{intersection of all } gHg^{-1}, \text{ is a normal subgroup of } G.$

Solution. By hypothesis $W = \bigcap_{g \in G} gHg^{-1}.$

Since each gHg^{-1} is a subgroup of G , W is a subgroup of G .

$$w \in W \Rightarrow w \in gHg^{-1} \forall g \in G \Rightarrow g^{-1}wg \in H \forall g \in G.$$

Let $x \in G$, then $xwx^{-1} = xg(g^{-1}wg)g^{-1}x^{-1} \in xgH(xg)^{-1} \forall g \in G.$

Now any element t of G can be written as $xx^{-1}t = xt$ where $g = x^{-1}t \in G$

$$\text{So } xwx^{-1} \in tHt^{-1} \forall t \in G \Rightarrow xwx^{-1} \in W$$

Hence W is a normal subgroup of G .

PROBLEMS

1. If f is a homomorphism of a group G into G' , prove that $G/\text{Ker } f \cong f(G).$

2. For any homomorphism $f: G \rightarrow H$ prove that, if $x \in G$ is of finite order, then $o(f(x)) \mid o(x)$ and conclude that f is a monomorphism (1-1) if and only if $o(x) = [o(f(x))] \forall x \in G.$

3. & 4. Classify the mappings given below into the following categories:

- (i) Non endomorphisms.
- (ii) Endomorphisms but not automorphisms.
- (iii) Isomorphisms but not automorphisms.
- (iv) Automorphisms.
- (v) Monomorphisms but not isomorphisms.

Let $\mathbb{Z}$ be the additive group of integers

$$(a) f_1: \mathbb{Z} \rightarrow \mathbb{Z}, f_1(x) = -x \forall x \in \mathbb{Z}.$$

$$(b) f_2: \mathbb{Z} \rightarrow \mathbb{Z}, f_2(x) = |x| \forall x \in \mathbb{Z}.$$

$$(c) f_3: \mathbb{Z} \rightarrow \mathbb{Z}, f_3(x) = x+2 \forall x \in \mathbb{Z}.$$

$$(d) f_4: \mathbb{Z} \rightarrow \mathbb{Z}, f_4(x) = -3x \forall x \in \mathbb{Z}.$$

Let C^* be the multiplicative group of all non-zero complex numbers.

$$(e) g_1: C^* \rightarrow C^*; g_1(x) = x^4 \forall x \in C^*.$$

$$(f) g_2: C^* \rightarrow C^*; g_2(x) = \bar{x} \forall x \in C^*.$$

$$(g) g_3: C^* \rightarrow C^*; g_3(x) = |x| \forall x \in C^*.$$

$$(h) g_4: C^* \rightarrow C^*; g_4(x) = -x \forall x \in C^*.$$

Let G be any abelian group of order 8.

$$(i) h_1: G \rightarrow G; h_1(x) = x^4 \forall x \in G.$$

$$(j) h_2: G \rightarrow G; h_2(x) = x^3 \forall x \in G.$$

$$(k) h_3: G \rightarrow G; h_3(x) = x^{-1} \forall x \in G.$$

5. Prove that a group G is abelian if and only if the mapping $f: G \rightarrow G$ given by $f(x) = x^{-1}$ is an automorphism of G .

6. Let G be an abelian group and n , a positive integer. Define $f: G \rightarrow G$ by $f(x) = x^n \forall x \in G$. Show that f is an endomorphism of G . If G is finite such that n and $o(G)$ are co-prime, show that f is an automorphism of G .

7. Let $f: G \rightarrow G'$ be a homomorphism. Prove the following:

(i) For any subgroup H of G , $f(H)$ is a subgroup of G' .

(ii) For any subgroup K' of G' , $f^{-1}(K')$ is a subgroup of G containing $\text{Ker } f$ and $f^{-1}(K')$ is normal in G whenever K' is normal in G' .

If f is onto then for any normal subgroup K of G , $f(K)$ is a normal subgroup of G' .

(iv) If f is onto and K is a subgroup of G containing $\text{Ker } f$ then $f^{-1}[f(K)] = K$ and conclude that there exists one-to-one correspondence between subgroups of G containing $\text{Ker } f = H$ and the subgroups of G' . This correspondence also gives a one-to-one correspondence between normal subgroups of G containing $\text{Ker } f$ and the normal subgroups of G' .

8. Prove that a homomorphic image of a cyclic (Abelian) group is cyclic (Abelian). Give an example to show that if a homomorphic image of a group is cyclic or Abelian, the same may not hold for the group itself.

9. Prove that if for a group G , $f: G \rightarrow G$ given by $f(x) = x^3$, $x \in G$, is an isomorphism, then G is Abelian.

10. Exhibit a non-Abelian group G , in which the map $f: G \rightarrow G$ given by $f(x) = x^4$ is an automorphism.

11. N is a normal subgroup of finite index in a group G and H is a subgroup of finite order. If $[G : N]$ is relatively prime to $o(H)$, show that $H \subseteq N$.

12. N is a finite normal subgroup of a group G and H is a subgroup of finite index in G . If $[G : H]$ is relatively prime to $o(N)$, then $N \subseteq H$. Prove!

13. If H is a subgroup of a group G , such that $x^2 \in H$ for all $x \in G$. Prove that G/H is abelian.

14. Let G and H be groups and f , a relation in $G \times H$ such that $\text{dom } f = G$. Further f satisfies the following conditions.

(i) $(x, a) \in f, (y, b) \in f \Rightarrow (xy^{-1}, ab^{-1}) \in f$.

(ii) If e, e' are the identities of G and H respectively, then $(e, a) \in f \Rightarrow a = e'$.

Prove that f is a homomorphism from G to H .

[Hint. Prove that $(x, a) \in f, (x, c) \in f \Rightarrow a = c$.]

15. Let G_i be a group for all $i = 1, 2, \dots, n$. Define $\pi G_i = G_1 \times G_2 \times \dots \times G_n$. Show that πG_i is a group under the composition

$$(a_1, a_2, \dots, a_n) (b_1, b_2, \dots, b_n) = (a_1 b_1, a_2 b_2, \dots, a_n b_n).$$

If H_i is a normal subgroup of G_i for all $i=1, 2, \dots, n$: prove that

$$\frac{\pi G_i}{\pi H_i} \cong \pi \left(\frac{G_i}{H_i} \right)$$

3. Permutation Groups

Let S be a set and let $A(S)$ denote the set of all permutations of S . We have proved earlier that $A(S)$ is a group. If H is any non-void subset of $A(S)$, such that H is closed under the resultant composition in $A(S)$ and for each element of H , its inverse also belongs to H , then H is a subgroup of $A(S)$. In other words one can say that if H is a set of permutations of a set S such that H is closed under resultant composition and inverse of each element of H is a member of H , then H is a group with respect to the resultant composition. Any such group is called a *transformation group* or a *permutation group*. These groups are of great significance since we will show in Cayley's Theorem (given ahead) that each group can be described in terms of a permutation group and further these groups are the richest source of different types of examples and counter-examples of groups with various properties. We shall mainly confine here to permutation groups acting on finite sets. Let S be a finite set having n elements and let $f \in A(S)$, we denote f by the symbol

$$\begin{pmatrix} a_1, a_2, \dots, a_n \\ b_1, b_2, \dots, b_n \end{pmatrix}$$

In the first row $a_1, a_2, \dots, a_n$ are the n elements of S , put in certain order. In the second row $b_1, b_2, \dots, b_n$ are such that for each a_i , b_i is the image of a_i under f and b_i is put just under a_i . It is immaterial in which order the elements of S are put in the first row. This notation is known as *2-rowed notation* for f .

EXAMPLE 9. Let $S = \{a_1, a_2, a_3, a_4\}$. Define $f : S \rightarrow S$ as follows :

$f(a_1) = a_2, f(a_2) = a_3, f(a_3) = a_1, f(a_4) = a_4$. Then we can write

$$\begin{aligned} f &= \begin{pmatrix} a_1 & a_2 & a_3 & a_4 \\ a_2 & a_3 & a_1 & a_4 \end{pmatrix} = \begin{pmatrix} a_1 & a_3 & a_4 & a_2 \\ a_2 & a_1 & a_4 & a_3 \end{pmatrix} \\ &= \begin{pmatrix} a_2 & a_3 & a_1 & a_4 \\ a_3 & a_1 & a_2 & a_4 \end{pmatrix}. \end{aligned}$$

Here we have indicated above only three ways of writing f . We can actually represent f in $4! = 24$ ways.

Coming back to general discussion, let elements of S be put in some definite order say $a_1, a_2, \dots, a_n$. Now $b_i = f(a_i)$. Since b_i is one of $a_1, a_2, \dots, a_n$, $b_i = a_{k_i}$ for some integer k_i lying between 1 and n . Since $b_1, \dots, b_n$ are all distinct, integers $k_1, k_2, \dots, k_n$ are all

the integers from 1 to n in some order. Hence

$$\begin{pmatrix} 1 & 2 & 3 & \dots & n \\ k_1 & k_2 & k_3 & \dots & k_n \end{pmatrix}$$

is a permutation of $\{1, 2, 3, \dots, n\}$. This permutation is naturally determined by f . Conversely given a permutation

$$\begin{pmatrix} 1 & 2 & 3 & \dots & n \\ k_1 & k_2 & k_3 & \dots & k_n \end{pmatrix}$$

of $\{1, 2, 3, \dots, n\}$ we get a permutation

$$\begin{pmatrix} a_1 & a_2 & \dots & a_n \\ a_{k_1} & a_{k_2} & \dots & a_{k_n} \end{pmatrix}$$

of S . Thus essentially any permutation f of S is uniquely determined by a permutation of $\{1, 2, \dots, n\}$ and symbol π does not play any part in the determination of f . So for convenience sake

$$\text{in place of } f = \begin{pmatrix} a_1 & a_2 & \dots & a_n \\ a_{k_1} & a_{k_2} & \dots & a_{k_n} \end{pmatrix}$$

we shall simply write

$$f = \begin{pmatrix} 1 & 2 & 3 & \dots & n \\ k_1 & k_2 & k_3 & \dots & k_n \end{pmatrix}.$$

Further we shall usually take $S = \{1, 2, 3, \dots, n\}$.

EXAMPLE 10. Let S and f be same as in Example 9. Then

$$f = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 3 & 1 & 4 \end{pmatrix}$$

Lemma 3.15. Let f be a permutation of a set S . Define a relation $\sim$ on S as follows: For any $a, b \in S$, $a \sim b$ if and only if $f^n(a) = b$ for some integer n . This relation is an equivalence relation.

Proof: Let $a, b, c \in S$.

(1) *Reflexivity.* Since $f^0(a) = I(a) = a$, $a \sim a$.

Here I is the identity permutation of S .

(2) *Symmetry.* $a \sim b \Rightarrow f^n(a) = b$ for some integer n

$$\Rightarrow f^{-n}(b) = a \Rightarrow b \sim a.$$

(3) *Transitivity.* $a \sim b, b \sim c \Rightarrow f^n(a) = b, f^m(b) = c$ for some integers m, n .

This

$$\begin{aligned} \Rightarrow f^{m+n}(a) &= (f^m \circ f^n)(a) = f^m(f^n(a)) \\ &= f^m(b) = c \Rightarrow a \sim c. \end{aligned}$$

This proves that ' $\sim$ ' is an equivalence relation on S . ■

Relation ' $\sim$ ' divides S into disjoint equivalence classes. Each equivalence class determined by the relation ' $\sim$ ' is called an *f-orbit* of S .

REMARKS. 1. When f is fixed, we drop f and write *orbit* in place of *f-orbit*.

2. If $x \in S$ is in *f-orbit* O of S we write O_x and read it as

f -orbit of x .

Lemma 3.16. If S is a finite set, $f \in A(S)$ and $s \in S$, then there exists a positive integer k , such that the f -orbit to which s belongs is, $\{s, f(s), f^2(s), \dots, f^{k-1}(s)\}$.

Proof: Since S is finite, $A(S)$ is of finite order; consequently $o(f)$ is finite. Let $o(f) = m$. Then $f^m = I \Rightarrow f^m(s) = I(s) = s$. Let k be the smallest positive integer such that $f^k(s) = s$. Then we claim $s = f^0(s), f(s), f^2(s), \dots, f^{k-1}(s)$ are all distinct. For, otherwise, we can find two integers l_1, l_2 satisfying $0 \leq l_1 < l_2 \leq k-1$ such that

$$f^{l_1}(s) = f^{l_2}(s).$$

Then $f^{l_2-l_1}(s) = s$ and $0 < l_2 - l_1 < k$.

This violates the minimality of k . Hence our assertion is true.

Now if any s' is in the orbit to which s belongs, then $s' = f^l(s)$ for some integer l . But $l = kq + r$ for some integers q, r such that $0 \leq r < k-1$.

Then $s' = f^l(s) = f^r[(f^k)^q(s)] = f^r(s)$, since $f^k(s) = s$ implies $f^{kq}(s) = s$ (why?). Further as $0 \leq r < k-1$, $f^r(s)$ is one of the elements $s, f(s), f^2(s), \dots, f^{k-1}(s)$. Hence $s, f(s), \dots, f^{k-1}(s)$ are the only distinct elements in the orbit to which s belongs. This proves the Lemma. ■

Definition 3.17. (Cyclic Permutation). Let S be a finite set. A permutation f of S is said to be a *cyclic permutation* or a *cycle* if there exist elements $i_1, i_2, \dots, i_t$ in S such that $f(i_1) = i_2, f(i_2) = i_3, \dots, f(i_{t-1}) = i_t, f(i_t) = i_1$ and for any $j \in S$ different from $i_1, i_2, \dots, i_t$, $f(j) = j$. We denote f by the symbol $(i_1 i_2 \dots i_t)$. This notation of f is called a *one-row notation*. Further t is called the *length* of the cycle f . A cycle of length t is also called a t cycle.

REMARK. Essentially one can say that a permutation f of a set S (arbitrary) is cyclic if S has at most one f -orbit having more than one element (justify!).

EXAMPLE 11. Let $S = \{1, 2, 3, 4\}$. Then (132) denotes the permutation f of S such that $f(1) = 3, f(3) = 2, f(2) = 1, f(4) = 4$. Thus in the two-rowed notation of f , we have

$$(132) = f = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 1 & 2 & 4 \end{pmatrix}.$$

f is a cyclic permutation of length 3. Notice further that $(132) = (213) = (321)$.

Similarly (12) is a cyclic permutation $\{1, 2, 3, 4\}$ of length two and

$$(12) = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 1 & 3 & 4 \end{pmatrix}.$$

EXAMPLE 12. Let $S = \{1, 2, 3, 4, 5\}$ and let

$$f = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 2 & 1 & 4 & 3 & 5 \end{pmatrix}.$$

We determine the orbits of S . Notice $f(1)=2, f(2)=1$. By applying Lemma 3.16 we see that $\{1, 2\}$ is an orbit of S . Again $f(3)=4, f(4)=3$; so that $\{3, 4\}$ is also an orbit of S . Since $f(5)=5, \{5\}$ is itself an orbit. It will be good for the readers to verify the following:

(i) The orbits of S under the permutation

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 2 & 3 & 1 & 5 & 4 \end{pmatrix}$$

are $\{1, 2, 3\}, \{4, 5\}$.

(ii) The only orbit of S under the permutation

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 2 & 3 & 5 & 1 & 4 \end{pmatrix}$$

is the set S itself.

Definition 3.18. (Transposition). A cycle of length 2 is called a Transposition.

Definition 3.19. (Disjoint permutation).

Two permutations f and g of a set S are said to be *disjoint* if they satisfy the following:

(i) For any $j \in S, f(j) \neq j \Rightarrow g(j) = j$

and (ii) For any $j \in S, g(j) \neq j \Rightarrow f(j) = j,$

i.e., if any element of S is moved by f then it is left fixed by g and if any element of S is moved by g then it is left fixed by f .

REMARK. Any two disjoint permutations f and g commute. For this let $x \in S$. Suppose $f(x) \neq x$, then $g(x) = x$. Now if $y = f(x)$ we see that $f(y) = f\{f(x)\} \neq y$, otherwise $x = f(x)$; thus $g(y) = y$. Hence $(gof)(x) = g\{f(x)\} = g(y) = y$. Also $(fog)(x) = f\{g(x)\} = f(x) = y$. Therefore $(fog)(x) = (gof)(x)$. Now let $f(x) = x$ then two cases arise, either $g(x) \neq x$ or $g(x) = x$.

In the former case, $g\{g(x)\} \neq g(x)$ so $f[g(x)] = g(x)$; further $gf(x) = g(x)$. Hence $fg(x) = gf(x)$.

In the latter case, $f\{g(x)\} = f(x) = x$ and $g\{f(x)\} = g(x) = x$. So once again we get $fg(x) = gf(x)$. Hence $fog = gof$ and this shows that g and f commute.

EXAMPLE 13. Let $S = \{1, 2, 3, 4, 5\}$. The cycle (12) is a transposition. Similarly $(23), (14)$ are also transpositions of S . Find all possible transpositions of S (they are 10 in number).

EXAMPLE 14. Let $S = \{1, 2, 3, 4, 5\}, f = (132)$ and $g = (45)$.

By Definition $f(1)=3, f(3)=2, f(2)=1$ but $g(1)=1, g(3)=3, g(2)=2$. Again $g(4)=5, g(5)=4$ but $f(4)=4, f(5)=5$. This shows that f and g are disjoint permutations.

REMARK. An easy way to check whether two non-identity cyclic

permutations are disjoint or not, is to see if they have any common element in their one-row notation or not. If they have any common element in their one-row notation then they are not disjoint otherwise they are disjoint. For example (123), (354) are not disjoint since 3 is common element in (123) and (354); but (12), (37) are disjoint since they have no element in common. By keeping in mind the above rule try to see which of the following pairs of cycles are disjoint (i) (12), (32); (ii) (452), (13); (iii) (135), (243). The justification of this remark is left to the readers.

Definition 3.20. (*Cycles of Permutation*). Let S be any finite non-void set and let f be any permutation of S , then for a given $s \in S$, as proved in Lemma 3.16, there exists a positive integer k such that $s, f(s), f^2(s), \dots, f^{k-1}(s)$ are all distinct and $\{s, f(s), \dots, f^{k-1}(s)\}$ is an f -orbit of S . The cyclic permutation $(sf(s) \dots f^{k-1}(s))$ is called a cycle of f .

Thus for a permutation f , if $\sigma = (sf(s) \dots f^{k-1}(s))$ is a cycle of f then for any s' equal to any of $s, f(s), \dots, f^{k-1}(s)$; $\sigma(s') = f(s')$.

EXAMPLE 15. Let $S = \{1, 2, 3, 4, 5\}$ and

$$f = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 2 & 3 & 1 & 5 & 4 \end{pmatrix}.$$

Then $(1f(1)f^2(1)) = (123)$ is a cycle of f . Again as $f(4) = 5, f^2(4) = f\{f(4)\} = f(5) = 4$, $(4, 5)$ is also a cycle of f . In fact (123) and (45) are only cycles of f . Notice that these two cycles are disjoint.

Lemma 3.21. Any two cycles of a permutation of a finite set are disjoint.

Proof: Let $(s_1 f(s_1) \dots f^{k-1}(s_1))$ and $(s_2 f(s_2) \dots f^{l-1}(s_2))$ be any two distinct cycles of f . By definition of a cycle we know that the sets $\{s_1, f(s_1), \dots, f^{k-1}(s_1)\}$ and $\{s_2, f(s_2), \dots, f^{l-1}(s_2)\}$ are two distinct f -orbits of S . Now any two distinct orbits of S do not have any element in common (Lemma 3.15). So $(s_1 f(s_1) \dots f^{k-1}(s_1))$ and $(s_2 f(s_2) \dots f^{l-1}(s_2))$ have no element in common. This implies that the two cycles are disjoint. ■

We are now ready to prove an important theorem.

Theorem 3.21. Any non-identity permutation of finite set can be expressed uniquely as a product of disjoint cycles, each of length > 1 , except for the order in which the cycles are written.

Proof: Let S be any non-void finite set and $f \in A(S)$. Let $\sigma_1, \sigma_2, \dots, \sigma_t$ be the totality of cycles of f . These cycles are pairwise disjoint (Lemma 3.21), we show that $f = \sigma_1 \sigma_2 \dots \sigma_t$.

Let

$$\sigma_1 = (a_1 a_2 \dots a_r)$$

$$\sigma_2 = (b_1 b_2 \dots b_s)$$

$$\dots \dots \dots$$

$$\sigma_i = (l_1 l_2 \dots l_u).$$

Consider $\sigma_1 \sigma_2 \dots \sigma_t = (a_1 a_2 \dots a_r)(b_1 b_2 \dots b_s) \dots (l_1 l_2 \dots l_u)$. Let $x \in S$, since each f -orbit of S determines a unique cycle of f and each element of S occurs in some f -orbit of S , there exists a cycle σ_i such that x is an element of σ_i .

Let $\sigma_i = (k_1 k_2 \dots k_m)$ and $x = k$, for some $1 \leq j \leq m$. Then by definition $f(x) = k_{j+1}$ if $j < m$ and $f(x) = k_1$ if $j = m$ i.e. $f(x) = \sigma_i(x)$. Since no element of σ_i occurs in any of the cycles $\sigma_1, \sigma_2, \dots, \sigma_{i-1}, \sigma_{i+1}, \dots, \sigma_t$, these cycles leave every $k_1, k_2, \dots, k_m$ fixed. Consequently $\sigma_1 \sigma_2 \dots \sigma_t(x) = \sigma_i(x) = f(x)$. Hence $f = \sigma_1 \sigma_2 \dots \sigma_t$. Note that upto now we have not used the fact that f is a non-identity permutation on S and also we have not shown that length of each cycle σ_i is greater than 1. As a matter of convention, in any decomposition of f as a product of cycles, cycles of length 1 are dropped.

$$\begin{aligned} \text{Let } f &= \sigma_1 \sigma_2 \dots \sigma_v. & \dots (1) \\ &= n_1 n_2 \dots n_w. & \dots (2) \end{aligned}$$

be two expressions of f as a product of disjoint cycles each of length > 1 . Consider any $n_l (l = 1, 2, \dots, w)$, since n_l is a cycle we have $n_l = (c_1 c_2 \dots c_m)$ for some $c_1, c_2, \dots, c_m$ in $\{1, 2, \dots, n\}$. Then $c_k (k = 1, 2, \dots, m)$ are not in any n_i with $i \neq l$. So $n_i(c_k) = c_k \forall i \neq l$ and we have $f(c_k) = n_l(c_k) = c_{k+1} \forall k = 1, 2, \dots, m-1$ and $f(c_m) = n_l(c_m) = c_1$. $\dots (3)$

As $f(c_1) = n_l(c_1) = c_2, f^2(c_1) = f(c_2) = n_l(c_2) = c_3, f^3(c_1) = f(c_3) = n_l(c_3) = c_4, \dots, f^{m-1}(c_1) = c_m$ and $f^m(c_1) = c_1; \{c_1, c_2, \dots, c_m\}$ is an orbit of f . Now if c_1 were not an element of any σ_j then this would have given that $\sigma_j(c_1) = c_1 \forall j = 1, 2, \dots, v$ and hence $f(c_1) = c_1$; which is not possible since $f(c_1) = c_2 \neq c_1$. So there exists σ_i such that c_1 is an element of σ_i . As for n_l , we can also prove that the elements of σ_i constitute an f -orbit of S . Since c_1 is an element of σ_i we see that the orbit $\{c_1, c_2, \dots, c_m\}$ constitutes all the elements of σ_i . Thus, as for n_l , we also have $\sigma_r(c_k) = c_k$ and $f(c_k) = \sigma_i(c_k) \forall k = 1, 2, \dots, m$ and $r \neq i$. Consequently, using (3), we get $c_2 = f(c_1) = \sigma_i(c_1), c_3 = f(c_2) = \sigma_i(c_2), \dots, c_m = f(c_{m-1}) = \sigma_i(c_{m-1})$ and $c_1 = f(c_m) = \sigma_i(c_m)$. Thus we see that $\sigma_i(c_1) = c_1, \sigma_i(c_2) = c_3, \dots, \sigma_i(c_{m-1}) = c_m, \sigma_i(c_m) = c_1$. Hence $\sigma_i = (c_1 c_2 \dots c_m) = n_l$. Thus each n_l is equal to some σ_i . Since every σ_r with $r \neq i$ is disjoint from σ_i and hence not equal to σ_i , we get n_l is equal to σ_i for unique i . Similarly each σ_k is equal to some unique n_j . Thus there is a 1-1 correspondence between σ_i 's and n_i 's such that corresponding cycles are equal. This establishes the uniqueness. ■

Corollary 3.23. Every permutation of a finite set S having more than one element, is a product of transposition.

Proof: Consider any cycle $(a_1 a_2 \dots a_m)$ of length m . If $m > 1$, clearly $(a_1 a_2 \dots a_m) = (a_1 a_m)(a_1 a_{m-1}) \dots (a_1 a_3)(a_1 a_2)$. If $m=1$, then, as S contains more than one element, $\exists b \in S$ such that $b \neq a_1$, we get that $(a_1) = (a_1 b)(b a_1)$. By the above theorem, each permutation is a product of cycles. Hence we can express each permutation as a product of transpositions. ■

Theorem 3.24. *Let f be a permutation of a finite set S , then in all expressions of f as product of transpositions either the number of transpositions is always even or always odd.*

Proof: Let $f = \sigma_1 \sigma_2 \dots \sigma_t$ be the expressions of f as a product of disjoint cycles, each of length > 1 . If some element, say x , of S is not in any of the cycle σ_i then we add (x) i.e. the cycle of length 1 in the above expression of f .

In this manner each element of S is in some cycle of f . Let now $f = \sigma'_1 \sigma'_2 \dots \sigma'_k$ where σ'_i are disjoint cycles, each of length ≥ 1 . We

put $N(f) = \sum_{i=1}^k [l(\sigma'_i) - 1]$, where $l(\sigma')$ denotes the length of the cycle σ' . Since for each cycle σ' of length 1, $l(\sigma') - 1 = 0$, $N(f)$

$= \sum_{i=1}^t [l(\sigma_i) - 1]$. $N(f)$ is uniquely defined non-negative integer. We

show that any factorization of f as a product of disjoint cycles has even or odd number of transpositions according as the $N(f)$ is even or odd. This will help us in proving the result. For this purpose we need the following formulae:

$$(ab)(ac_1 c_2 \dots c_h b d_1 d_2 \dots d_k) = (b d_1 d_2 \dots d_k)(a c_1 c_2 \dots c_h)$$

$$\text{and } (ab)(b d_1 \dots d_k)(a c_1 c_2 \dots c_h) = (a c_1 c_2 \dots c_h b d_1 d_2 \dots d_k).$$

According to these formulae if a and b occur in the same cycle of f then $N[(ab)f] = N(f) - 1$ and if a and b occur in different cycles of f then $N[(ab)f] = N(f) + 1$.

$$\text{If any case } N[(ab)f] = N(f) \pm 1 \quad \dots (1)$$

We now suppose that f is a product of m transpositions, say

$$f = (a_1 b_1)(a_2 b_2) \dots (a_m b_m).$$

We prove the result by induction on m . If $m=1$, then $N(f)=1$ and the result holds trivially.

Suppose that $m > 1$ and that the result holds for every permutation expressible as a product of $m-1$ transpositions.

$$\text{Now } (a_1 b_1)f = (a_2 b_2) \dots (a_m b_m).$$

By induction hypothesis $N[(a_1 b_1)f]$ and $m-1$ are both even or both odd. However by (1), $N[(a_1 b_1)f] = N(f) \pm 1$. This shows $N(f)$ and m are either both odd or both even. Hence the proof is completed. ■

Note: Another proof of this theorem is given in worked-out exercises after Theorem 3.34.

Definition 3.25. (Even or Odd Permutation). A permutation f of a finite non-void set S is said to be *even* or *odd* according as f is expressible as a product of even or odd number of transpositions.

Theorem 3.26. For any $n > 1$, the subset A_n of S_n consisting of all even permutations is a normal subgroup of S_n of index 2.

Proof: Let $X = \{1, 2, \dots, n\}$ and let S_n be the group of all permutations of X . Since $I = (12)(12)$, $I \in A_n$. Thus A_n is non-void. Now let $f, g \in A_n$ and

$$f = (a_1 b_1) (a_2 b_2) \dots (a_t b_t) \text{ and } g = (c_1 d_1) (c_2 d_2) \dots (c_u d_u)$$

be expressions of f and g as product of even number of transpositions; so that both t and u are even. Consequently $t+u$ is even.

However $fog^{-1} = (a_1 b_1) \dots (a_t b_t) (c_u d_u) (c_{u-1} d_{u-1}) \dots (c_1 d_1)$ is a product of $t+u$ permutations, we conclude that fog^{-1} is an even permutation. Therefore $f \in A_n, g \in A_n \Rightarrow fog^{-1} \in A_n$.

Hence A_n is a subgroup.

Consider $(12)(12)$ is an odd permutation so that $(12) \notin A_n$. Consequently $A_n(12) \neq A_n$. Since every even permutation in S_n is in A_n , if we show that every odd permutation of S_n is in $A_n(12)$, then it will follow that every member of S_n is either in A_n or in $A_n(12)$. That will give us $S_n = A_n \cup A_n(12)$ and A_n has index two in S_n . Since every subgroup of index two is normal, we shall also get that A_n is normal in S_n . So let $f \in S_n$ and f be odd. Clearly then $f(12)$ is even; so that $f(12) \in A_n$. Thus $f = [f(12)](12) \in A_n(12)$. Hence the result follows.■

Definition 3.26 (a). For any $n \geq 1$, S_n is called the *symmetric group* of degree n .

Definition 3.26 (b). For any $n \geq 1$, A_n is called the *alternating group* of degree n .

Theorem 3.27. (Cayley) Every group is isomorphic to a permutation group.

Proof: Let G be a group and let $A(G)$ denote the group of all permutations of the set G . For each $a \in G$ define $f_a : G \rightarrow G$ by $f_a(x) = ax \forall x \in G$. For any $x, y \in G$, $f_a(x) = f_a(y) \Rightarrow ax = ay \Rightarrow x = y$. So f_a is 1-1. Further $f_a(a^{-1}x) = a(a^{-1}x) = x$ gives that f_a is also onto. Hence $f_a \in A(G)$. Now for any $a, b, x \in G$, $(f_a \circ f_b)(x) = f_a[f_b(x)] = f_a(bx) = a(bx) = (ab)x = f_{ab}(x)$; this gives that $f_a \circ f_b = f_{ab}$. Define $\sigma : G \rightarrow A(G)$ by $\sigma(a) = f_a \forall a \in G$.

Then for all $a, b \in G$, $\sigma(ab) = f_{ab} = f_a \circ f_b = \sigma(a) \circ \sigma(b)$.

Moreover $\sigma(a) = \sigma(b) \Rightarrow f_a = f_b \Rightarrow f_a(e) = f_b(e) \Rightarrow ae = be \Rightarrow a = b$.

Thus σ is a 1-1 homomorphism of G into $A(G)$. Hence G is iso-

morphic to $\sigma(G)$ which, being a subgroup of $A(G)$, is a permutation group. This proves the theorem.

The group of permutations S_3 on the set $X = \{1, 2, 3\}$ plays important role in group theory. It is the smallest non-commutative group. We shall show that there are only two types of groups of order six.

Theorem 3.28. There are only two groups of order six, one is cyclic and other is isomorphic to S_3 .

Proof: Let G be a group of order 6. Since there are five non-identity elements and $x \leftrightarrow x^{-1}$ is a 1-1 correspondence between the elements and their inverses, there exists $a (\neq e) \in G$ such that a is its own inverse i.e. $o(a) = 2$. Let G be Abelian and H be the sub-group $\langle a \rangle$. Now $o(G/H) = 3$ implies that G/H is cyclic. Let bH be a generator of G/H . Since $o(bH) \mid o(b)$, we get $o(b) = 3$ or 6. If $o(b) = 6$ then trivially G is cyclic. If $o(b) = 3$ then $o(ab) = 6$ as $ab = ba$ and $o(a)$ and $o(b)$ are coprime (Lemma 4.23). Hence again G is cyclic group generated by ab . Suppose now G is not cyclic then G cannot be Abelian. Then all nonidentity elements of G cannot be of order 2 (Chapter 2, § 2, Problem 10), thus there exists c in G such that $o(c) = 3$ and $ac \neq ca$. Let $H = \{e, c, c^2\}$. As $[G : H] = 2$, H is a normal subgroup. Consequently $a^{-1}ca \in H$. So that $a^{-1}ca = c$ or c^2 ; but as $ac \neq ca$, we get $a^{-1}ca = c^2$; i.e. $ca = ac^2$. Now $a \notin H$ so we get $G = H \cup aH = \{e, c, c^2, a, ac, ac^2\}$. It is a routine verification, that the mapping $f : G \rightarrow S_3$ given by $f(e) = I, f(a) = (12), f(c) = (123), f(c^2) = (132), f(ac) = (23)$ and $f(ac^2) = (13)$ is an isomorphism.

Hence the result follows. ■

With the help of Theorem 3.28 we can show that converse of Lagrange's theorem is not true. This we do in Theorem 3.31. Firstly we prove a lemma.

Lemma 3.29. S_4 has no element of order ≥ 4 .

Proof: S_4 acts upon $\{1, 2, 3, 4\}$. Let $\sigma (\neq I) \in S_4$ then $\sigma = \sigma_1 \sigma_2 \dots \sigma_t$, where $\sigma_1, \sigma_2, \dots, \sigma_t$ are pairwise disjoint cycles each of length > 1 . Since the symbols which occur in σ_i are among $\{1, 2, 3, 4\}$, no σ_i can be of length > 4 , and no two σ_i, σ_j ($i \neq j$) have any symbol in common, we have $t \leq 2$. Thus either σ is cyclic or σ can be expressed as a product of two disjoint cycles each of length two. If σ is cyclic then it is of length 2, 3 or 4; hence σ is of order 2, 3 or 4. If σ is a product of two disjoint cycles, each of length 2, then σ is of order 2. Hence S_4 has no element of order exceeding 4. ■

Theorem 3.30. A_4 has no subgroup of order six.

Proof: By the above lemma A_4 has no element of order 6. So

that A_4 can't have a cyclic subgroup of order 6. Thus if at all there is a subgroup of A_4 of order 6 it must be isomorphic to S_3 (Theorem 3.28). Let H be such a subgroup of A_4 . As $H \cong S_3$, H contains three elements of order 2 each and two elements of order 3 each. The only elements of A_4 of order 2 are $(12)(34)$, $(13)(24)$ and $(14)(23)$. Let σ be an isomorphism between H and S_3 . Now $\sigma[(12)(34)(13)(24)] = \sigma[(14)(23)]$. $(14)(23)$ is of order 2 $\Rightarrow$ order of $\sigma[(14)(23)] = 2$. But in S_3 , product of two different elements of order 2 is an element of order 3. Thus in S_3 order of $\sigma[(14)(23)] = 3$, which is a contradiction. This proves that our assumption is wrong. Hence A_4 has no subgroup of order 6. ■

Note. See a shorter proof of this theorem in worked out exercises, at the end of Theorem 3.34.

REMARK. Note that $o(A_4) = 12$ and $6 \mid 12$. Hence Theorem 3.30 shows that the converse of Lagrange's theorem is not true.

We recall that a group $G[\neq(e)]$ is called a *simple group* if it has no proper normal subgroups.

Every group of prime order is simple since it has no proper subgroups what to talk of proper normal subgroups! Naturally the question arises whether there are simple groups of composite orders? The answer is, yes; indeed there are infinitely many simple groups of composite orders. To show this it is sufficient to prove that A_n is simple for all $n \geq 5$. We shall prove this result with the help of some lemmas. In all of these lemmas we always take $n \geq 5$.

Lemma 3.31. *Every element of A_n is a product of 3 cycles.*

Proof: Each $x \in A_n$ is expressible as a product of even number of transpositions and each transposition $(\alpha\beta)$ where $\alpha \neq 1$, $\beta \neq 1$ can be expressed as $(1\alpha)(1\beta)(1\alpha)$. So x itself can be written as $(1\alpha_1)(1\alpha_2) \dots (1\alpha_r)$ where r is even and $\alpha_i \in \{2, 3, \dots, n\}$. Further $(1\alpha)(1\beta) = (1\beta\alpha)$ and r is even $\Rightarrow$ every even permutation is a product of 3-cycles. ■

REMARK. It may be noted that we required $n \geq 3$ (not $n \geq 5$) in the above proof.

Lemma 3.32. *If a normal subgroup H of A_n contains a cycle of length 3, then $H = A_n$.*

Proof: In view of Lemma 3.31, it is sufficient to prove that every cycle of length 3 belongs to H .

Let $(\alpha\beta\gamma)$ be any cycle of length 3 and let $(xyz) \in H$. Consider $\theta \in S_n$ such that $\theta(x) = \alpha$, $\theta(y) = \beta$ and $\theta(z) = \gamma$.

Then $\theta(xyz)\theta^{-1} = (\alpha\beta\gamma)$; thus if $\theta \in A_n$, then $(\alpha\beta\gamma) \in H$, as H is a normal subgroup of A_n .

So let $\theta \notin A_n$. Take u, v in $\{1, 2, \dots, n\}$ different from x, y, z

(this is possible as $n \geq 5$). As θ is odd, $\theta(uv) \in A_n$.

So that $\theta(uv)(xyz)[\theta(uv)]^{-1} \in H$

$$\Rightarrow \theta(uv)(xyz)(uv)\theta^{-1} \in H$$

$\Rightarrow \theta(xyz)\theta^{-1} \in H$, as (xyz) and (uv) are disjoint and hence $(uv)(xyz)(uv) = (xyz)(uv)^2 = (xyz)$.

$$\Rightarrow (\alpha\beta\gamma) \in H.$$

Hence the lemma follows. ■

Lemma 3.33. *If any normal subgroup H of A_n contains a product of two disjoint transpositions, then $H = A_n$.*

Proof: Let $(xy)(uv) \in H$ where (xy) and (uv) are disjoint.

Take w in $\{1, 2, 3, \dots, n\}$ different from x, y, u and v ; this is possible as $n \geq 5$.

Put $\theta = (uvw)$. Clearly $\theta \in A_n$, so that $\theta(xy)(uv)\theta^{-1} \in H$ i.e. $(uvw)(xy)(uv)(wvu) \in H$

$$\text{i.e. } (vw)(xy) \in H.$$

However H is a subgroup, therefore $(xy)(uv)(vw)(xy) \in H$

$$\text{i.e. } (uvw) \in H.$$

Hence $H = A_n$ (Lemma 3.32). ■

Now we are ready to prove our main result.

Theorem 3.34. *(Galois) A_n is simple for $n \geq 5$.*

Proof: Let $H \neq \{I\}$ be a normal subgroup of A_n . We shall show that $H = A_n$. Since $H \neq \{I\}$, $\exists \alpha (\neq I) \in H$. Let $\alpha = \sigma_1 \sigma_2 \dots \sigma_k$ be the decomposition of α into disjoint cycles, each of length > 1 . Since disjoint cycles commute we can assume that length of $\sigma_i \geq$ length of $\sigma_{i+1} \forall i = 1, 2, \dots, k-1$.

Four cases arise

(i) At least one of σ_i is of length > 3 .

(ii) σ_1 and σ_2 are of length 3.

(iii) σ_1 is of length 3 and all other σ_i 's are transpositions.

(iv) All σ_i 's are transpositions.

Case I. Since there is at least one cycle of length > 3 and according to our choice lengths are in descending order of magnitude σ_1 is a cycle of length > 3 .

$$\text{Let } \sigma_1 = (a_1 a_2 \dots a_r), \quad r > 3.$$

Consider $\theta = (a_1 a_2 a_3)$. Clearly $\theta \in A_n$, so that $\theta \alpha \theta^{-1} \in H$ (H is normal in A_n);

$$\text{i.e., } (a_1 a_2 a_3) \sigma_1 \sigma_2 \dots \sigma_k (a_3 a_2 a_1) \in H$$

$$\Rightarrow \sigma_2 \sigma_3 \dots \sigma_k (a_1 a_2 a_3) (a_1 a_2 \dots a_r) (a_3 a_2 a_1) \in H, \text{ as } \sigma_i \text{'s are disjoint,}$$

$$\Rightarrow \sigma_2 \sigma_3 \dots \sigma_k (a_1 a_4 a_5 \dots a_r a_2 a_3) \in H$$

But H is a subgroup, therefore

$$\alpha^{-1} \sigma_2 \sigma_3 \dots \sigma_k (a_1 a_4 a_5 \dots a_r a_2 a_3) \in H$$

$$\Rightarrow \sigma_k^{-1} \sigma_{k-1}^{-1} \dots \sigma_2^{-1} \sigma_2 \sigma_3 \dots \sigma_k \sigma_1^{-1} (a_1 a_4 a_5 \dots a_r a_2 a_3) \in H, \text{ as } \sigma_i \text{'s}$$

are disjoint

$$\Rightarrow (a_r a_{r-1} \dots a_3 a_2 a_1)(a_1 a_4 a_5 \dots a_r a_2 a_3) \in H$$

$$\Rightarrow (a_1 a_3 a_r) \in H.$$

Consequently $H = A_n$ (Lemma 3.32).

Case II. Let $\sigma_1 = (a_1 a_2 a_3)$ and $\sigma_2 = (a_4 a_5 a_6)$

Consider $\theta = (a_5 a_6 a_1)$

Since $\theta \in A_n$ and H is a normal subgroup of A_n .

$$\theta \alpha \theta^{-1} \in H;$$

$$i.e. \quad (a_5 a_6 a_1) \sigma_1 \sigma_2 \dots \sigma_k (a_1 a_6 a_5) \in H$$

$$\Rightarrow \sigma_3 \sigma_4 \dots \sigma_k (a_5 a_6 a_1) (a_1 a_2 a_3) (a_4 a_5 a_6) (a_1 a_6 a_5) \in H$$

$$\Rightarrow \sigma_3 \sigma_4 \dots \sigma_k (a_1 a_4 a_6) (a_2 a_3 a_5) \in H.$$

This gives $\alpha^{-1} \sigma_3 \sigma_4 \dots \sigma_k (a_1 a_4 a_6) (a_2 a_3 a_5) \in H$

$$i.e. \quad \sigma_k^{-1} \sigma_{k-1}^{-1} \dots \sigma_2^{-1} \sigma_1^{-1} \sigma_3 \sigma_4 \dots \sigma_k (a_1 a_4 a_6) (a_2 a_3 a_5) \in H$$

$$\Rightarrow (\sigma_k^{-1} \sigma_{k-1}^{-1} \dots \sigma_3^{-1} \sigma_3 \sigma_4 \dots \sigma_k) \sigma_2^{-1} \sigma_1^{-1} (a_1 a_4 a_6) (a_2 a_3 a_5) \in H$$

$$\Rightarrow (a_6 a_5 a_1) (a_3 a_2 a_1) (a_1 a_4 a_6) (a_2 a_3 a_5) \in H$$

$$\Rightarrow (a_1 a_6 a_3 a_4 a_5) \in H.$$

and we are led to case (I). Thus, once again $H = A_n$.

Case III. Let $\sigma_1 = (a_1 a_2 a_3)$ and each σ_i for $i=2, 3, \dots, k$ be transpositions,

$$\alpha = \sigma_1 \sigma_2 \dots \sigma_k \in H$$

$$\Rightarrow \alpha^2 = \sigma_1^2 \sigma_2^2 \dots \sigma_k^2 \in H$$

$$\Rightarrow \sigma_1^2 \in H, \text{ as } \sigma_i^2 = 1 \text{ for all } i=2, 3, \dots, k.$$

$$\Rightarrow (a_1 a_2 a_3)^2 \in H$$

$$\Rightarrow (a_3 a_2 a_1) \in H.$$

Consequently $H = A_n$ (Lemma 3.32).

Case IV. Let $\sigma_1 = (a_1 a_2)$ and $\sigma_2 = (a_3 a_4)$.

Consider $\theta = (a_3 a_4 a_1)$. Because $\theta \in A_n$ and H is normal subgroup of A_n , $\theta \alpha \theta^{-1} \in H$

$$i.e. \quad (a_3 a_4 a_1) \sigma_1 \sigma_2 \dots \sigma_k (a_1 a_4 a_3) \in H$$

$$\Rightarrow \sigma_3 \sigma_4 \dots \sigma_k (a_3 a_4 a_1) (a_1 a_2) (a_3 a_4) (a_1 a_4 a_3) \in H$$

$$\Rightarrow \sigma_3 \sigma_4 \dots \sigma_k (a_1 a_4) (a_2 a_3) \in H.$$

Now H is a subgroup, therefore

$$\alpha^{-1} \sigma_3 \sigma_4 \dots \sigma_k (a_1 a_4) (a_2 a_3) \in H$$

$$\Rightarrow \sigma_k^{-1} \sigma_{k-1}^{-1} \dots \sigma_3^{-1} \sigma_2^{-1} \sigma_1^{-1} \sigma_3 \sigma_4 \dots \sigma_k (a_1 a_4) (a_2 a_3) \in H$$

$$\Rightarrow (a_4 a_3) (a_2 a_1) (a_1 a_4) (a_2 a_3) \in H$$

$$\Rightarrow (a_1 a_3) (a_2 a_4) \in H.$$

This implies that $H = A_n$ (Lemma 3.33).

Hence A_n is a simple group for every $n \geq 5$. ■

REMARK. Trivially A_n is simple for $n=3$. So we can say that A_n is simple for all natural numbers n except 1, 2 and 4.

WORKED OUT EXERCISES

Exercise 1. Prove that in S_n the number of distinct cycles of length $r \leq n$ is $\frac{1}{r} \frac{n!}{(n-r)!}$.

Solution. The number of distinct arrangements of r numbers chosen out of n numbers, is equal to ${}^nP_r = \frac{n!}{(n-r)!}$.

As each of the cycles $(a_1 a_2 \dots a_r)$, $(a_2 a_3 \dots a_r a_1)$, $\dots$, $(a_r a_1 a_2 \dots a_{r-1})$ is same, the number of distinct r -cycles is equal to $\frac{1}{r} \frac{n!}{(n-r)!}$.

Exercise 2. A_4 has no subgroup of order six.

Solution. Since a 3-cycle of S_4 is an even permutation, all 3-cycles of S_4 belong to A_4 . By Exercise 1 above, the number of distinct 3-cycles in A_4 is equal to $\frac{1}{3} \frac{4!}{(4-3)!} = 8$.

Let H be a subgroup of A_4 , of order 6. All 3-cycles of A_4 can not belong to H . Let x be a 3-cycle of A_4 such that $x \notin H$. As $o(x)=3$, the subgroup $K=\{e, x, x^2\}$ consists of three elements, where e is the identity of A_4 .

Now $x \notin H$, $x^2 = x^{-1} \notin H \Rightarrow H \cap K = \{e\} \Rightarrow o(HK) = \frac{o(H)o(K)}{o(H \cap K)} = \frac{6 \cdot 3}{1} = 18$.

This is absurd as HK is a subset of A_4 and A_4 has 12 elements. Hence our supposition is wrong. Consequently A_4 has no subgroup of order six.

Exercise 3. Let f be a permutation of a finite S , then in all expressions of f as product of transpositions either the number of transpositions is always even or always odd.

Solution. Let the result be not true. We can find a permutation $f \in A(S)$ which can be expressed as a product of an even number of transpositions and at same time, as a product of an odd number of transpositions.

Thus $f = (a_1 b_1)(a_2 b_2) \dots (a_r b_r)$, $a_i, b_i \in S$,
 $= (c_1 d_1)(c_2 d_2) \dots (c_s d_s)$, $c_i, d_i \in S$,

where r is an even number and s is an odd number.

Then $f^{-1} = (c_s d_s) \dots (c_2 d_2)(c_1 d_1)$, as $(xy)^{-1} = (xy)$ for all $x, y \in S$. If I is the identity permutation, $I = ff^{-1} \Rightarrow I = (a_1 b_1) \dots (a_r b_r)(c_s d_s) \dots (c_1 d_1)$ is a decomposition of I into a product of $r+s$, i.e. an odd number of transpositions. In this way we can write $I = (\alpha_1 \beta_1) \dots (\alpha_k \beta_k)$ with $\alpha_i, \beta_i \in S$ and k , an odd number.

Consider $(\alpha_i \beta_i)$. If $\alpha_i \neq 1, \beta_i \neq 1$, we can write $(\alpha_i \beta_i) = (1 \beta_i)(1 \alpha_i)(1 \beta_i)$.

As a consequence $I = (1r_1)(1r_2) \dots (1r_t)$ where t is an odd number and $r_1, r_2, \dots, r_t \in S - \{1\}$.
 Consider the effect of I on r_t . $I(r_t) = r_t$ (1)

Further $(1r_t)$ takes r_t to 1, so in order to bring it back to r_t , $(1r_t)$ must be repeated, in left hand side of (1), an even number of times.

Similar arguments can be given to show that each transposition $(1r_j)$ $1 \leq j \leq t$, must be repeated an even number of times. This implies that t must be an even number. Thus we get a contradiction. Hence our assumption is wrong and so the statement of the problem holds true.

Exercise 4. H is a subgroup of S_n for some integer $n > 1$. If H contains an odd permutation, show that the set of all even permutations in H form a normal subgroup of H of index 2.

Solution. Put $X = \{1, -1\}$, X is a subgroup of real numbers under multiplication.

Define $\theta: H \rightarrow X$ by $\theta(f) = 1$ if f is even permutation in H and $\theta(f) = -1$ if f is an odd permutation in H . It can be verified that θ is a homomorphism. It is onto, as H contains an odd permutation. Also $\text{Ker } \theta = \text{set of all even permutations of } H$. As $\text{Ker } \theta$ is a normal subgroup of H , the first part of the problem follows.

By Fundamental Theorem of Homomorphism $\frac{H}{\text{Ker } \theta} \cong X \Rightarrow \frac{o(H)}{o(\text{Ker } \theta)} = o(X) = 2 \Rightarrow \text{Ker } \theta$ is of index 2 in H .

Exercise 5. If H is a subgroup of a group G and S is the set of all left cosets of H in G , then there is a homomorphism θ of G into $A(S)$ such that $\text{Ker } \theta$ is the largest normal subgroup of G which is contained in H .

Solution. Define $\theta: G \rightarrow A(S)$ by $\theta(g) = T_g$ where $T_g(xH) = gxH$ for all $xH \in S$. Firstly, we show that $T_g \in A(S)$. Clearly T_g is a mapping from S to S . Also $T_g(xH) = T_g(yH) \Rightarrow gxH = gyH \Rightarrow (gy)^{-1}gx \in H \Rightarrow y^{-1}x \in H \Rightarrow xH = yH \Rightarrow T_g$ is 1-1. Further any left coset $xH \in S$ can be written as $g(g^{-1}xH)$ hence T_g is onto. Consequently $T_g \in A(S)$. Again $\theta(gh) = T_{gh}$ where $T_{gh}(xH) = (gh)xH = g(hxH) = T_g(hxH) = T_gT_h(xH) \Rightarrow T_{gh} = T_gT_h$. Hence $\theta(gh) = T_gT_h = \theta(g)\theta(h)$. This in turn implies that θ is a homomorphism from G to $A(S)$.

Now $g \in \text{Ker } \theta \Rightarrow T_g = I$ where I is the identity of $A(S)$. This gives that $T_g(xH) = xH \forall xH \in S \Rightarrow gxH = xH \forall xH \in S \Rightarrow geH = eH \Rightarrow gH = H \Rightarrow g \in H$. Hence $\text{Ker } \theta \subseteq H$. Further if N a normal subgroup of G , contained in H ; then for each $n \in N$, $\theta(n) = T_n$ where $T_n(xH) = nxH = x(x^{-1}nx)H$ for all $x \in G$. But

N is a normal subgroup of G , $x^{-1}nx \in N \subseteq H \Rightarrow x^{-1}nx \in H$. So we set $T_n(xH) = xH$ for all $x \in G$ i.e. $T_n = I$. Hence $n \in \text{Ker } \theta$. In other words $N \subseteq \text{Ker } \theta$.

Consequently $\text{Ker } \theta$ is the largest normal subgroup of G contained in H .

Exercise 6. If G is a finite group and $H \neq G$ is a subgroup of G such that $o(G) \nmid i(H)!$, then H must contain a non-trivial normal subgroup of G .

Solution. By Exercise 5, $\text{Ker } \theta \subseteq H$. Since $H \neq G$, $\text{Ker } \theta \neq G$. Further if $\text{Ker } \theta = \{e\}$ where e is the identity of G , then $\frac{G}{\text{Ker } \theta} \cong T$ where T is a subgroup of $A(S)$, gives that $o(G) = o(T)$ must be a factor of $o[A(S)]$. But $o[A(S)] = i(H)!$. This in turn implies that $o(G) \mid i(H)!$, which is against the hypothesis. Hence $\text{Ker } \theta \neq \{e\}$. Thus H contains a non-trivial normal subgroup of G .

Exercise 7. G is a group of order $2n$, where n is odd. Show that G has a normal subgroup of order n .

Solution. G possesses at least one element of order 2, otherwise we can partition G into union of disjoint pairs $\{a, a^{-1}\}$, $\{b, b^{-1}\}$, ... etc and a singleton $\{e\}$, where e is the identity of G . This will in turn imply that number of elements in G is odd! So, let $a \in G$ be of order 2. Consider now T_a , where $T_a(x) = ax$ for all $x \in G$. T_a is a permutation on the set G . (Verify!). Since $T_a^2(x) = T_a[T_a(x)] = a(ax) = a^2x = ex = x$, each T_a -orbit of G consists of two elements and number of distinct orbits of $T_a = 2n/2 = n$. As each orbit corresponds to a cycle, T_a has n cycles each of length 2. Let $\sigma_1, \sigma_2, \dots, \sigma_n$ be all the n 2-cycles (i.e. transpositions) of T_a . So $T_a = \sigma_1\sigma_2 \dots \sigma_n$. As n is odd, T_a is an odd permutation in $A(G)$. Let $K = \{T_g \mid g \in G\}$. K is a subgroup of $A(G)$ and $G \cong K$ under the mapping $f(g) = T_g$. (Check!). As $T_a \in K$, K contains an odd permutation. So by Exercise 4 we get that K contains a normal subgroup N of index 2. Thus G contains a normal subgroup $H = f^{-1}(N)$ of index 2. But $o(H) = \frac{o(G)}{i(H)} = \frac{2n}{2} = n$. Hence G has a normal subgroup of order n .

PROBLEMS

1. Find the orbits and cycles of the set $\{1, 2, 3, 4, 5, 6\}$ under following permutations and express each of them as a product of its cycles:

$$(i) \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 2 & 1 & 5 & 3 & 4 \end{pmatrix}$$

$$(ii) \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 3 & 4 & 2 & 5 & 1 & 6 \end{pmatrix}$$

$$(iii) \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 1 & 4 & 3 \end{pmatrix}$$

$$(iv) \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 4 & 2 & 1 \end{pmatrix}$$

2. Express the following permutations as products of disjoint cycles:

$$(i) (123)(45)(1345).$$

$$(ii) (12)(54)(32)(17)(28).$$

$$(iii) (45)(123)(321)(54)(26)(14).$$

3. Let f be a permutation of a finite set and let $\sigma_1 \sigma_2 \dots \sigma_r$ be the expression of f as a product of disjoint cycles. If $o(\sigma_i) = m_i \forall i$, show that $o(f) = \text{lcm of } m_1, m_2, \dots \& m_r$.

4. Let $f = (a_1 a_2 \dots a_n)$ be a cycle of length $n (> 1)$. Prove the following:

$$(i) f^{r-1}(a_1) = a_r, 2 \leq r \leq n.$$

$$(ii) \text{ Conclude from (i) that } o(f) \geq n.$$

$$(iii) \text{ Show that } f^n(a_i) = a_i, 1 \leq i \leq n.$$

$$(iv) \text{ Prove that } o(f) = n.$$

5. Show that a cycle of length n is an odd or an even permutation according as n is even or odd.

6. Using the fact that any non-void finite closed subset of a group is a subgroup, prove the following:

$$(i) H = \{I, (12)(34), (13)(24), (14)(23)\} \text{ is a normal subgroup of } A_4.$$

$$(ii) K = \{I, (14)(23)\} \text{ is a normal subgroup of } H.$$

$$(iii) K \text{ is a subgroup of } A_4, \text{ but } K \text{ is not normal in } A_4.$$

(Thus $K \subset H \subset A_4$, K is normal in H , H is normal in A_4 , however, K is not normal in A_4).

7. Prove the following:

(i) Every permutation f in S_n ($n > 1$) is expressible as a product of transpositions $(12), (13), \dots, (1n)$.

(ii) Any permutation f in S_n ($n > 1$) is expressible as a product of $(12), (23), (34), \dots, (n-1, n)$.

[Hint. Use part (i)].

(iii) $H = \{I, (12)\}$ and $K = \{I, (123), (132)\}$ are two subgroups of S_3 such that neither of them is normal, but $KH = HK$. (This provides an example of two subgroups H and K , which are not normal but their product HK is a subgroup).

8. Consider $H = \{I, (12)\}$ and $K = \{I, (13)\}$, two subgroups of S_3 ; show that $HK \neq KH$.

9. Prove that a group of order p^2 , where p is a prime number, must have a normal subgroup of order p .

10. G is a group of order $2^k m$ where $k \geq 1$ is an integer and m is an odd number. If G has a cyclic subgroup of order 2^k then prove that G has a normal subgroup of index 2^k .

Now $a \in \text{Ker } g \Leftrightarrow g(a)=I$, the identity of $\text{In}(G)$

$$\Leftrightarrow f_a = I$$

$$\Leftrightarrow f_a(x)=I(x) \quad \forall x \in G$$

$$\Leftrightarrow axa^{-1}=x \quad (\text{Definition of } f_a)$$

$$\Leftrightarrow a \in Z(G) \quad [\text{Definition of } Z(G)].$$

Hence $\text{Ker } g = Z(G)$ and this completes the proof. ■

REMARK. If a group G is not Abelian then $Z(G) \neq G$ and consequently $G/Z(G)$ is not a trivial group; further because of the above theorem $\text{In}(G)$ is also not a trivial group. Consequently if a group G is not Abelian, then G has a non-trivial automorphism. If a group G is Abelian, then each of its inner automorphism is identity, so that above theorem need not provide us a non-identity automorphism of an Abelian group.

Lemma 3.39. Any subgroup of a cyclic group is cyclic.

Proof: Let $G = \langle a \rangle$ be a cyclic group and let H be a subgroup of G . If $H = (e)$, then trivially H is cyclic. Suppose that $H \neq (e)$ so there exists $a^n \in H$ such that $a^n \neq e$. Then $a^{-n} = (a^n)^{-1} \in H$. As either n or $-n$ is a positive integer, we can say that $a^n \in H$ for some positive integer n .

Let k be the least positive integer such that $a^k \in H$. If we show that $H = \langle a^k \rangle$, then it will follow that H is a cyclic group. For this let $b \in H$, as $b \in G$, $b = a^m$ for some integer m . The division algorithm gives $m = kq + r$ for some integers q, r with $0 \leq r < k$. Now $a^r = a^m(a^k)^{-q} \in H$. The minimality of k implies $r = 0$. Thus $m = kq$ and $b = (a^k)^q$. Hence $H = \langle a^k \rangle$. This proves the lemma. ■

Corollary 3.40. Any subgroup of the additive group of integers is of the form $\langle n \rangle$, where n is some non-negative integer.

Proof: Since the additive group of integers is cyclic the result follows from the above lemma. ■

Theorem 3.41. Any infinite cyclic group is isomorphic to the additive group of integers. Any finite cyclic group of order n is isomorphic to the quotient group $\mathbf{Z}/\langle n \rangle$.

Proof: Let $G = \langle a \rangle$ be a cyclic group with a as its generator. Define $f: \mathbf{Z} \rightarrow G$ by $f(n) = a^n \quad \forall n \in \mathbf{Z}$. Given $n, m \in \mathbf{Z}$, $f(n+m) = a^{n+m} = a^n a^m = f(n)f(m)$. So f is a homomorphism. Further as each member of G is of the type a^n which is equal to $f(n)$, f is also onto.

Case I. If $G = \langle a \rangle$ is of infinite order then $a^n \neq e \quad \forall n \neq 0$; so that $\text{Ker } f = (0)$. Hence f is an isomorphism of $\mathbf{Z}$ onto G .

Case II. Let $o(G) = n$ then $o(a) = n$. Thus for any integer m , $a^m = e \Leftrightarrow n \mid m \Leftrightarrow m \in \langle n \rangle$. Hence $\text{Ker } f = \langle n \rangle$. By the Fundamental Theorem of Homomorphism, $\frac{\mathbf{Z}}{\text{Ker } f} \cong G$. This gives that

$\frac{\mathbb{Z}}{\langle n \rangle} \cong G$. This proves the theorem. ■

Since $\mathbb{Z} \cong \mathbb{Z}/\langle 0 \rangle$, if G is an infinite cyclic group we can say that $G \cong \mathbb{Z}/\langle 0 \rangle$. So we combine case I and case II in the proof of above theorem and get:

Theorem 3.42. Any cyclic group is isomorphic to $\mathbb{Z}/\langle n \rangle$ for some non-negative integer n . ■

As the relation of isomorphism of groups is an equivalence relations. Theorem 3.42 gives the following.

Theorem 3.43. Any two cyclic groups of same order are isomorphic. ■

Let us recall that any element a of a cyclic group G is called a generator of G if each element of G is expressible as a power of a . For example consider the group $G = \{1, -1, i, -i\}$ where $i = \sqrt{-1}$ and $-i$ are generators of G but 1 and -1 are not generators.

Theorem 3.44. Any infinite cyclic group has two and only two generators. If $G = \langle a \rangle$ is a finite cyclic group of order n then a^m , for any positive integer m less than n , is a generator of G if and only if the highest common factor of m and n is equal to 1.

Proof: Let $G = \langle a \rangle$ be an infinite cyclic group. Suppose $a^n \in G$ is also a generator of G . Since $a \in G$, by definition of a generator $a = (a^n)^m$ for some integer m . That means $a^{nm} = a = a^1$. As a is of infinite order we get $nm = 1$. Hence either $n = 1, m = 1$ or $n = -1, m = -1$, i.e. $a^n = a$ or a^{-1} . Thus possibly a^{-1} is the only generator of G different from a . But, given any $a^k \in G, a^k = (a^{-1})^{-k}$ i.e., every element of G is expressible as a power of a^{-1} . Consequently a^{-1} is also a generator of G . Hence a and a^{-1} are the only generators of G .

Now let $G = \langle a \rangle$ be of finite order n . Then $G = \{a^0 = e, a^1, a^2, \dots, a^{n-1}\}$. If $n = 1$ then, $G = \langle e \rangle$ so the result holds trivially. Let $n > 1$ and a^m ($1 \leq m < n$) be any generator of G . Then for some integer $k, a = (a^m)^k \Rightarrow a^{mk-1} = e \Rightarrow n \mid (mk-1) \Rightarrow mk-1 = ln$ for some integer l ; thus $mk - ln = 1$. This gives that HCF of m and n i.e. $(m, n) = 1$. Conversely, suppose that m is a positive integer $< n$ such that $(m, n) = 1$. Then there exist integers x and y such that $mx + ny = 1$. Thus $a = a^1 = a^{mx+ny} = (a^m)^x (a^n)^y = (a^m)^x$ as $a^n = e$. For any integer $k, a^k = (a^m)^{xk}$. This shows that every element of G is a power of a^m . Hence a^m generates G . Hence the theorem is proved. ■

Definition 3.45. (Euler's ϕ -function)

For any positive integer $n, \phi(n)$ denotes the number of those positive integers which are less than n and are relatively prime to n . $\phi(n)$ is called Euler's totient function.

REMARK. For some integer $n > 1$, let $G = \langle a \rangle$ be a cyclic group of order n . Since $e = a^0, a^1, \dots, a^{n-1}$ are all distinct and for any posi-

11. For $a \in A_n$, define $\pi_a : A_n \rightarrow A_n$ by $\pi_a(x) = axa$ for all $x \in A_n$. Let $C(A_n) = \{\pi_a \mid a \in A_n\}$. Show that $o(A_n) = o[C(A_n)]$ for $n \geq 4$.
12. If S and T are two finite sets such that $o(T) = o(S) = n$, then show that $S_n \cong T_n$.
13. Show $Z(S_3)$ i.e. the centre of S_3 , consists of identity permutation only.

14. Show that the group $S = \bigcup_{n=1}^{\infty} S_n$ contains exactly one normal subgroup A distinct from S and $\{e\}$, where e is the identity of S . Further show that A is of index 2 in S .

15. Show that S_3 is a homomorphic image of S_4 .
16. If a permutation f on n symbols can be factored into k disjoint cycles, then show that the number of transpositions in any representation of f as a product of transpositions is either always even or always odd according as $n-k$ is even or odd.

4. Automorphisms of Groups and structure of Cyclic Groups

Let G be a group. $A(G)$ denotes the set of all permutations in G . We know that $A(G)$ is a group under the resultant composition. Let $Aut(G)$ denote the set of all automorphisms of G . From the definition, we know that each automorphism of G is a one to one mapping of G onto itself. Thus $Aut(G) \subseteq A(G)$. We also know that the identity map is an automorphism. However the problem of determination of all possible automorphisms of a group in general is fairly difficult. Still we can prove the following:

Theorem 3.35. For any group G , the set $Aut(G)$ of all automorphisms of G is a group under the resultant composition.

Proof: Since identity map I on G is in $Aut(G)$, $Aut G \neq \phi$.

(I) *Closure:* Let $f, g \in Aut(G)$. Then, as f, g are both 1-1 mappings of G onto G , fog is also a 1-1 mapping of G onto G . Further for any $a, b \in G$, $(fog)(ab) = f[g(ab)] = f[g(a)g(b)]$, since g is an automorphism. As f is also an automorphism and $g(a), g(b) \in G$, we have $f[g(a)g(b)] = [f\{g(a)\}][f\{g(b)\}] = [fog(a)][fog(b)]$.

i.e. $(fog)(ab) = [(fog)(a)][(fog)(b)]$.

Consequently fog is an automorphism and $fog \in Aut(G)$. This proves the closure of $Aut(G)$.

(II) *Associativity:* Since the resultant composition for mapping is in general associative, the composition in $Aut(G)$ is also associative.

(III) *Existence of identity:* As $I \in Aut(G)$ and $foI = f = Iof \forall f \in Aut(G)$, I is the identity of $Aut(G)$.

(IV) *Existence of Inverses:* Let $f \in Aut(G)$. As f is 1-1 mapping of G onto G , f^{-1} is defined and f^{-1} is also a 1-1 mapping of G onto

G . Now let $a, b \in G$; since f is onto $\exists x, y \in G$ such that $f(x)=a$ and $f(y)=b$; this gives, $f^{-1}(a)=x$ and $f^{-1}(b)=y$. Consequently $ab=f(x)f(y)=f(xy)=f[f^{-1}(a)f^{-1}(b)]$. Hence $f^{-1}(ab)=f^{-1}(a)f^{-1}(b)$.

Thus we see that f^{-1} is a 1-1 endomorphism of G onto G . So that f^{-1} is an automorphism of G . Consequently $f^{-1} \in \text{Aut}(G)$ and it is such that $f \circ f^{-1} = f^{-1} \circ f = I$.

Hence $\text{Aut}(G)$ is a group. ■

REMARK. Alternatively the above result can be proved as follows:

$$f, g \in \text{Aut}(G) \Rightarrow f \circ g^{-1} \in \text{Aut}(G) (?)$$

Hence $\text{Aut}(G)$ is a subgroup of $A(G)$. This proves $\text{Aut}(G)$ is a group.

Lemma 3.36. Let G be a group and $a \in G$. Then the mapping $f_a: G \rightarrow G$ defined by $f_a(x) = axa^{-1}$ is an automorphism of G .

For proof see Example 8. ■

Definition 3.37. (Inner automorphism). Let a be an element of a group G . The automorphism $f_a: G \rightarrow G$ given by $f_a(x) = axa^{-1} \forall x \in G$ is called an inner automorphism of G determined by a .

NOTATION. $\text{In}(G)$ denotes the set of all inner automorphisms of G .

Theorem 3.38. For any group G , $\text{In}(G)$ is a normal subgroup of $\text{Aut}(G)$. Further $\text{In } G \cong G/Z(G)$, where $Z(G)$ denotes the centre of G .

Proof: Clearly $I \in \text{In}(G)$ as $I(x) = x = exe^{-1} = f_e(x) \forall x \in G$.

Now for any $a \in G, x \in G, f_a \circ f_{a^{-1}}(x) = f_a[f_{a^{-1}}(x)] = f_a[a^{-1}x(a^{-1})^{-1}] = f_a(a^{-1}xa) = a(a^{-1}xa)a^{-1} = x = I(x)$.

This implies

$$f_a \circ f_{a^{-1}} = I.$$

Similarly

$$(f_{a^{-1}} \circ f_a)(x) = I(x).$$

Thus $f_{a^{-1}} \circ f_a = I = f_a \circ f_{a^{-1}}$ showing thereby that $(f_a)^{-1} = f_{a^{-1}} \in \text{In}(G)$.

Also for any $f_a, f_b \in \text{In}(G)$ simple computation shows that $f_a \circ f_b = f_{ab} \in \text{In}(G)$. Hence $\text{In}(G)$ is a subgroup of $\text{Aut}(G)$.

To show that $\text{In}(G)$ is a normal subgroup of $\text{Aut}(G)$, it only remains to prove that for any $f_a \in \text{In}(G), \sigma \in \text{Aut}(G), \sigma \circ f_a \circ \sigma^{-1} \in \text{In}(G)$.

Let $x \in G$ then $(\sigma \circ f_a \circ \sigma^{-1})(x) = (\sigma \circ f_a)[\sigma^{-1}(x)] = \sigma[a\sigma^{-1}(x)a^{-1}] = \sigma(a)\sigma[\sigma^{-1}(x)]\sigma(a^{-1}) = \sigma(a)x[\sigma(a)]^{-1} = f_{\sigma(a)}(x)$.

Hence $\sigma \circ f_a \circ \sigma^{-1} = f_{\sigma(a)} \in \text{In}(G)$.

Now we consider the last part of the theorem. For this purpose define a mapping $g: G \rightarrow \text{In}(G)$ by $g(a) = f_a \forall a \in G$.

Then $g(ab) = f_{ab} = f_a \circ f_b = g(a) \circ g(b) \forall a, b \in G$ gives that g is a homomorphism. g is onto since each member of $\text{In}(G)$ is of the form f_a and by definition $f_a = g(a)$. Then by applying Fundamental Theorem of Homomorphism we get $\frac{G}{\text{Ker } g} \cong \text{In}(G)$. The result will follow if we show that $\text{Ker } g = Z(G)$.

tive integer $m < n$, a^m is generator of G if and only if $(m, n) = 1$. We get that number of generators of G is $\phi(n)$.

Theorem 3.46. Let G be a cyclic group of order n , and a positive integer m be a divisor of n . Then there exists one and only one subgroup of G of order m .

Proof: Let $G = \langle a \rangle$, $o(G) = o(a)$. Now $m \mid n \Rightarrow n = mq$ for some positive integer q .

Then $H = \langle a^q \rangle$ is a cyclic subgroup of G of order m . Let K be any other subgroup of G of order m . If k is the least positive integer such that $a^k \in K$, then $K = \langle a^k \rangle$ (Lemma 3.32) and for any integer s , $a^s \in K \Rightarrow k \mid s$. Now $a^n = e \in K = k \mid n$. Thus $o(a^k) = n/k$. Hence $o(K) = o(a^k) = n/k$. However $o(K) = m$. So $n = mk = mq$ i.e., $q = k$ and we get $K = \langle a^k \rangle = \langle a^q \rangle = H$. Hence $H = \langle a^q \rangle$ is the only subgroup of G of order n . ■

Theorem 3.47. Let $G \neq (e)$ be a cyclic group $\langle a \rangle$. Then any endomorphism f of G is an automorphism of G if and only if $f(a)$ is a generator of G . Further

(i) If G is infinite then $\text{Aut}(G)$ is of order 2.

(ii) If G is of finite order n , then $\text{Aut}(G)$ is isomorphic to the group of those positive integers $< n$, which are relatively prime to n , with binary composition as multiplication modulo n and order of $\text{Aut}(G)$ is $\phi(n)$.

Proof: Let f be an automorphism of G and let $x \in G$ then there exists an element $y \in G$, such that $x = f(y)$. But $G = \langle a \rangle$, so $y = a^n$ for some integer n , hence $x = f(y) = f(a^n) = \{f(a)\}^n$. Thus $G = \langle f(a) \rangle$. Conversely let $G = \langle f(a) \rangle$. Let $x \in G$, then $x = \{f(a)\}^k$ for some integer k . In other words $x = f(a^k)$, hence f is onto. Now if G is finite then f is clearly 1-1. Hence f is an automorphism. If G is infinite then $f(x) = f(y) \Rightarrow f(a^r) = f(a^s)$ where $x = a^r$, $y = a^s$ for some integers r and s as $G = \langle a \rangle$. Thus $\{f(a)\}^r = \{f(a)\}^s \Rightarrow \{f(a)\}^{s-r} = e$. But then G is infinite $\Rightarrow r = s$ i.e., $x = y$. Hence again f is 1-1. Consequently f is an automorphism, (i) now, follows from the fact that the number of generators of an infinite cyclic group is 2 (Theorem 3.44). Also if G is finite then the number of generators is $\phi(n)$ (Theorem 3.44). This implies that the order of $\text{Aut}(G)$ is $\phi(n)$. Let $f: G \rightarrow G$ be an automorphism, then $f(a) = a^k$. Since $f(a)$ is a generator of G , $k < n$ and $(k, n) = 1$ (Theorem 3.44). Define $\theta: \text{Aut } G \rightarrow Z_n$ where Z_n is the group of positive integers under multiplication modulo n , by $\theta(f) = \bar{k} \forall f \in \text{Aut}(G)$. θ is homomorphism since if $g: G \rightarrow G$ is another automorphism given by $g(a) = a^l$ then $(f \circ g)(a) = a^{kl} \Rightarrow \theta(fg) = \overline{kl} = \bar{k}\bar{l} = \theta(f)\theta(g)$. Also θ is 1-1, since $\theta(f) = \theta(g) \Rightarrow \bar{k} = \bar{l}$

$\Rightarrow n \mid k-l; k < n, l < n$ and $(k, n)=1, (l, n)=1 \Rightarrow k=l$. Thus $f=g$. Finally, since $o[\text{Aut}(G)]=\phi(n)$ and $o(Z_n)=\phi(n)$, θ is onto. This completes the proof of (ii). ■

Theorem 3.47. *Let G be a group with at least two elements such that G has no subgroup other than $\{e\}$ and itself. Then G is a cyclic group of prime order.*

Proof: First we claim that G is cyclic.

Let $a \in G, a \neq e$ (This choice is possible as G has at least two elements). Then $H = \langle a \rangle$ is a subgroup of G and $a \in H \Rightarrow H \neq \{e\}$. By hypothesis $H=G$. Hence $G = \langle a \rangle$.

Now we prove that G is of finite order. If not, then in view of Theorem 2.33, $o(a)$ can not be finite.

Consider $K = \langle a^2 \rangle$. K is a subgroup of G . $K = \{e\} \Rightarrow a^2 = e \Rightarrow a$ is of finite order, a contradiction. So $K=G$. But $a \in G \Rightarrow a \in \langle a^2 \rangle \Rightarrow a = (a^2)^i = a^{2i}$ for some $i \in \mathbb{Z}$. This gives $a^{2i-1} = e \Rightarrow a$ is of finite order, again a contradiction.

Finally suppose that G is of order n . If n is not prime, we can write $n=rs$ where $1 < r < n$. By Theorem 3.46, G has a subgroup H of order r . As $1 < r < n$, $H \neq \{e\}$ and $H \neq G$. This is against the hypothesis of theorem.

Consequently n is prime. ■

PROBLEMS

1. A subgroup H of a group G is said to be *characteristic subgroup* if for every automorphism f of G , $f(H) \subseteq H$.

(a) Prove the following:

(i) Let H be a normal subgroup of a group G . Show that any characteristic subgroup K of H is normal in G .

[Hint. Use the fact that every inner automorphism of G gives an automorphism of H .]

(ii) G is an abelian group and n is a fixed positive integer. Let $H = \{x \in G \mid x^n = e, \text{ where } e \text{ is the identity of } G\}$. Show that H is a characteristic subgroup of G .

(b) Give an example of a normal subgroup which is not a characteristic subgroup.

2. Let G be a finite group and T , a cyclic subgroup of G , by using the fact that for any divisor of the order of a finite cyclic group there exists one and only one subgroup of that order, prove that if T is normal in G then each subgroup of T is normal in G .

3. Let $G = \langle a \rangle$ and $G' = \langle b \rangle$ be any two cyclic groups of same order. Define $f: G \rightarrow G'$ by $f(a^r) = b^r$ for all integers r . Show

that f is well-defined and it is an isomorphism.

4. Let $G_1 = \langle a \rangle$, $G_2 = \langle b \rangle$ be two cyclic groups of order n and m respectively such that $(n, m) = 1$. Let $G = G_1 \times G_2 = \{x, y \mid x \in G_1, y \in G_2\}$. For any $(x_1, y_1), (x_2, y_2) \in G_1 \times G_2$ define $(x_1, y_1)(x_2, y_2) =$ that $(x_1 x_2, y_1 y_2)$. Show that G is a cyclic group of order mn .

[Hint. Show firstly that G is a group. Then, show that $(a, b) \in G$ is of order mn and $o(G) = mn$.]

5. Show that $\text{Aut } S_3 \cong S_3$.

6. Find out $\text{Aut } (G)$, where G is the Klein's 4-group.

(Ans. $\text{Aut } G \cong S_3$)

7. Show that two groups A and B are isomorphic if and only if there are two new operations $\circ$ and $*$ defined over the elements of groups as follows:

For all $a, a' \in A, b, b' \in B, a \circ b \in A, b * a \in B$.

Moreover $a \circ (b * a') = (a \circ b) a'$ and $b * (a \circ b') = (b * a) b'$.

8. Let G be a group and f , an endomorphism of G , which commutes with every inner automorphism of G . Let K be the set of elements $x \in G$ such that $\varphi^2(x) = \varphi(x)$. Show that K is a normal subgroup of G and G/K is Abelian.

9. Let G be a group, and $f: G \rightarrow G$ defined by $f(x) = x^n$ be an automorphism. Show that for each $a \in G, a^{n-1} \in Z(G)$, (i.e. centre of G).

10. Let T be an automorphism of a finite group G such that $T(x) = x$ for $x \in G$ if and only if $x = e$, also $T^2 = I$, the identity of $\text{Aut}(G)$. Show that G is Abelian.

[Hint. First prove that every $g \in G$ can be expressed as $x^{-1}T(x)$ for some $x \in G$.]

11. Let $M \subset N$ be two normal subgroups of a group G such that $o(N/M) = 2$ and G/N is cyclic. Show that G/M is Abelian.

Give an example of a group G with two normal subgroups $M \subset N$ such that both G/N and N/M are cyclic but G/M is not Abelian.

[Hint. Take $G = S_4, N = A_4, M = \text{Klein's 4-group}$.]

12. Let $\{G_i\}, i = 1, 2, \dots$, be subgroups of a group G such that $G_i < G_{i+1}$ for all i , show that $\bigcup G_i$ can't be a cyclic subgroup of G .

13. Show that the group $\mathbb{Q}$ of all rational numbers under addition is not cyclic.

14. Let G be a group and A , a group of automorphisms of G . Consider the set $S = A \times G$. For any $(f, a), (g, b) \in A \times G$ define $(f, a) * (g, b) = [fog, g^{-1}(a)b]$.

Prove the following:

(i) S is a group under $*$.

(ii) The mapping $\sigma: G \rightarrow S$ such that $\sigma(a) = (I, a) \forall a \in G$ is a

monomorphism.

$$(iii) \quad \begin{matrix} S \\ \sigma(G) \cong A. \end{matrix}$$

15. Show that if G is a group of order n such that for every positive divisor d of n , there is a unique subgroup of G of order d , then G is cyclic.

[Hint. For $d > 0$ dividing n , if $\psi(d)$ denotes the number of elements of order d , show that $\psi(d) = \phi(d)$. Also use the fact that $\sum_{d|n} \phi(d) = n$.]

5. Conjugate Elements

Definition 3.48. Let G be a group and $a, b \in G$. a is said to be a conjugate of b in G if there exists $c \in G$ such that $a = c^{-1}bc$.

Symbol $a \sim b$ will denote that a is a conjugate of b .

Lemma 3.49. The relation of conjugacy on a group G is an equivalence relation.

Proof: (1) Reflexivity : Since $a = a^{-1}aa$, $a \sim a$.

(2) Symmetry : $a \sim b \Rightarrow a = x^{-1}bx$ for some $x \in G$,
 $\Rightarrow b = xax^{-1} = y^{-1}ay$ where $y = x^{-1}$,
 $\Rightarrow b \sim a$.

(3) Transitivity : $a \sim b$, $b \sim c \Rightarrow a = x^{-1}bx$, $b = y^{-1}cy$
 for some $x, y \in G \Rightarrow a = x^{-1}(y^{-1}cy)x = (yx)^{-1}c(yx) \Rightarrow$
 $a \sim c$.

This completes the proof. ■

Above lemma shows that we can divide a group G into equivalence classes under the relation of conjugacy. Each such equivalence class is called a conjugate class. For any $a \in G$, $C(a)$ denotes the conjugate class to which a belongs i.e., $C(a) = \{c \in G \mid a \sim c\}$. Precisely $C(a)$ consists of all elements of the type $x^{-1}ax$. Therefore $a \in Z(G)$ if and only if $C(a) = \{a\}$. In particular $C(e) = \{e\}$.

Let G be a finite group. Let $C(a_1), C(a_2), \dots, C(a_t)$ be the totality of all conjugate classes of G ; let $C_{a_1}, C_{a_2}, \dots, C_{a_t}$ be their respective orders. Since they are pairwise disjoint and their union is G

we get $o(G) = \sum_{i=1}^t C_{a_i}$, or simply $o(G) = \sum_a C_a$ where the sum runs over

elements a , taken one each from each conjugate class. Henceforth in this article our task is to determine the nature of C_a and exhibit a few applications of the concept of conjugate elements.

Definition 3.50. (Centralizer or Normalizer of an element).

For any element a of a group G , the set $N(a) = \{x \in G \mid ax = xa\}$ is called the Normalizer or Centralizer of a in G .

One can see clearly from the definition of centre of a group that a is in the centre of G if and only if $N(a)=G$.

Lemma 3.51. For any $a \in G$, $N(a)$ is a subgroup of G .

Proof: If e is the identity of G then $ea=ae$ gives $e \in N(a)$. So that $N(a) \neq \phi$. Let $x \in N(a)$, $y \in N(a)$, then $(xy)a = x(ya) = x(ay) = (xa)y = (ax)y = a(xy)$; thus $xy \in N(a)$. Again $ax = xa \Rightarrow x^{-1}a = ax^{-1} \Rightarrow x^{-1} \in N(a)$. Hence $N(a)$ is a subgroup of G . ■

Theorem 3.52. If G is a finite group and $a \in G$ then $C_a = \frac{o(G)}{o[N(a)]}$.

Proof: Let $o(G)=n$. If $N(a)$ has t distinct right cosets: $N(a)x_1, N(a)x_2, \dots, N(a)x_t$, then we know that $t = \frac{o(G)}{o[N(a)]}$

Now for $1 \leq i, j \leq t$, $x_i^{-1}ax_i = x_j^{-1}ax_j$

$$\Rightarrow x_i x_j^{-1} a x_j x_i^{-1} = a \Rightarrow x_i x_j^{-1} a (x_i x_j^{-1})^{-1} = a$$

$$\Rightarrow (x_i x_j^{-1})a = a(x_i x_j^{-1}) \Rightarrow x_i x_j^{-1} \in N(a)$$

$$\Rightarrow N(a)x_i = N(a)x_j \Rightarrow i=j \text{ since } N(a)x_i \text{'s are all distinct.}$$

Hence $x_i^{-1}ax_i = x_j^{-1}ax_j \Rightarrow i=j$.

So $x_1^{-1}ax_1, x_2^{-1}ax_2, \dots, x_t^{-1}ax_t$ are all distinct conjugates of a . If we show that these are the only conjugates of a then it follows that

$C(a)$ contains only t elements i.e. $C_a = t = \frac{o(G)}{o[N(a)]}$.

Consider for some $x \in G$, $b = x^{-1}ax$. Since $G = \bigcup_{i=1}^t N(a)x_i$, $x = cx_i$ for some $c \in N(a)$ and some positive integer i . Therefore $x^{-1}ax = (cx_i)^{-1}a(cx_i) = x_i^{-1}(c^{-1}ac)x_i = x_i^{-1}ax_i$ since $c^{-1}ac = a$.

Hence any conjugate b of a is equal to one of $x_i^{-1}ax_i$. This proves that a has only t conjugates viz. $x_i^{-1}ax_i$, $i=1, 2, \dots, t$. Hence the theorem is proved. ■

Corollary 3.53. For any finite group G , $o(G) = \sum_a \frac{o(G)}{o[N(a)]}$ where the sum runs over elements a , taken one each from each conjugate class

Proof: Let $C(a_1), C(a_2), \dots, C(a_k)$ be the totality of conjugate classes of G . Then the fact these classes are pairwise disjoint and their union is G , gives $o(G) = \sum_{i=1}^k C_{a_i}$ where $C_{a_i} = o[C(a_i)]$.

However by the above theorem $C_{a_i} = \frac{o(G)}{o[N(a_i)]}$.

Hence $o(G) = \sum_{i=1}^k \frac{o(G)}{o[N(a_i)]}$. This proves the corollary ■.

Corollary 3.54. Let G be a finite group and $Z(G)$, its centre, then $o(G) = o[Z(G)] + \sum_a \frac{o(G)}{o[N(a)]}$ where the sum runs over elements a , taken one from each of those distinct conjugate classes which contain more than one element.

Proof: Since $a \in Z(G)$ if and only if $C_a = 1$, there are $o[Z(G)]$ number of conjugate classes each having only one element. Therefore the corollary follows. ■

Definition 3.55. Let G be a finite group. The equation $o(G) = o[Z(G)] + \sum_a \frac{o(G)}{o[N(a)]}$ where the sum runs over elements a , taken one from each of those distinct conjugate classes which contain more than one element, is called the *class equation* of the group G .

We wish to make a passing remark that the class equation has played great role in the development of the structure theory of non-commutative finite groups. Interested readers may find how often it is used in next chapter. Before we finish this section we prove two results on groups by using the class equation.

Theorem 3.56. If $o(G) = p^n$ where p is a prime number, then $Z(G) \neq (e)$.

Proof: Let $o[Z(G)] = z$. By the class equation $p^n = z + \sum_a \frac{o(G)}{o[N(a)]}$. . . (1)

where sum runs over elements a , taken one from each of conjugate classes $C(a)$ which has more than one element [i.e., $a \notin Z(G)$]. Now for each $a \notin Z(G)$, $o[N(a)] < o(G) = p^n$ and $o[N(a)] \mid o(G)$ gives

$$o[N(a)] = p^{n_a} \text{ for some } 1 \leq n_a < n \Rightarrow p \mid \frac{o(G)}{o[N(a)]} \quad \text{Further } p \mid p^n.$$

Hence from (1), we get $p \mid z$. This proves that

$$p \mid o[Z(G)] \Rightarrow o[Z(G)] > 1 \Rightarrow Z(G) \neq (e). \blacksquare$$

Corollary 3.57. If $o(G) = p^2$ where p is a prime, then G is Abelian.

Proof: As, a group G is Abelian if and only if $Z(G) = G$, it is sufficient to show that $o(G) = p^2 \Rightarrow Z(G) = G$. Let $o(G) = p^2$. By Lagrange's theorem $o[Z(G)] \mid p^2$. Therefore $o[Z(G)] = 1, p$ or p^2 . By Theorem 3.56, $o[Z(G)] \neq 1$, so $o[Z(G)] = p$ or p^2 . Suppose $o[Z(G)] = p$. Consider $a \in G$ such that $a \notin Z(G)$. Since for every $b \in Z(G)$, $ba = ab$, $b \in N(a)$. Thus $Z(G) \subseteq N(a)$. Also $a \in N(a)$ but $a \notin Z(G)$ so $N(a) \neq Z(G)$. Consequently $o[N(a)] > o[Z(G)] = p$. But $o[N(a)] \mid p^2$. Thus $o[N(a)] = p^2$ and $N(a) = G$. This implies $a \in Z(G)$. This contradicts the supposition that $a \notin Z(G)$. Therefore $o[Z(G)] = p^2$ and hence $G = Z(G)$. Thus the proof is complete ■

WORKED-OUT EXERCISES

Exercise 1. Two permutations σ and η of a finite set S are said to be similar if there exists a 1-1 correspondence between the cycles of σ and η such that the corresponding cycles have same length.

Show that two permutations of a finite set S are similar if and only if they are conjugate.

Solution Let σ and η be conjugate. There exists $g \in S_n$ such that $\eta = g^{-1}\sigma g$.

Let $\sigma = (a_1 a_2 \dots a_t) (b_1 b_2 \dots b_n) \dots (c_1 c_2 \dots c_s)$ where $a_i, b_j, \dots, c_k \in S$ for all $1 \leq i \leq t, 1 \leq j \leq n, \dots, 1 \leq k \leq s$ and all the cycles on the right are pairwise disjoint and each element of S is in some cycle. Since g is an onto mapping we can find $e_1, e_2, \dots, e_t, f_1, f_2, \dots, f_n, \dots, l_1, l_2, \dots, l_s$ such that $g(e_i) = a_i \forall 1 \leq i \leq t, g(f_j) = b_j \forall 1 \leq j \leq n, \dots, g(l_k) = c_k \forall 1 \leq k \leq s$.

Then $\eta = g^{-1}\sigma g = (e_1 e_2 \dots e_t) (f_1 f_2 \dots f_n) \dots (l_1 l_2 \dots l_s)$. As g is 1-1, lengths of corresponding cycles [under the correspondence $(a_1 a_2 \dots a_t) \leftrightarrow (e_1 e_2 \dots e_t), (b_1 b_2 \dots b_n) \leftrightarrow (f_1 f_2 \dots f_n), \dots, (c_1 c_2 \dots c_s) \leftrightarrow (l_1 l_2 \dots l_s)$] are equal. Hence σ and η are similar.

Conversely let σ and η be similar permutations in S_m . We can write $\sigma = (a_1 a_2 \dots a_t) (b_1 b_2 \dots b_n) \dots (c_1 c_2 \dots c_s)$ and $\eta = (x_1 x_2 \dots x_t) (y_1 y_2 \dots y_n) \dots (z_1 z_2 \dots z_s)$, where $t, n, \dots, s$ are integers ≥ 1 such that $t+n+\dots+s=m$.

$$\text{Put } h = \begin{pmatrix} a_1, a_2, \dots, a_t, b_1, b_2, \dots, b_n, \dots, c_1, c_2, \dots, c_s \\ x_1, x_2, \dots, x_t, y_1, y_2, \dots, y_n, \dots, z_1, z_2, \dots, z_s \end{pmatrix}$$

It can be verified that h is a permutation in S_m . Since $(h^{-1}\eta h)(a_1) = h^{-1}\eta(x_1) = h^{-1}(x_2) = a_2, \dots, (h^{-1}\eta h)(a_t) = h^{-1}\eta(x_t) = h^{-1}(x_1) = a_1, h^{-1}\eta h(b_1) = h^{-1}\eta(y_1) = h^{-1}(y_2) = b_2, \dots, h^{-1}\eta h(b_n) = h^{-1}\eta(y_n) = h^{-1}(y_1) = b_1, \dots, h^{-1}\eta h(c_1) = h^{-1}\eta(z_1) = h^{-1}(z_2) = c_2, \dots, h^{-1}\eta h(c_s) = h^{-1}\eta(z_s) = h^{-1}(z_1) = c_1$; we get $h^{-1}\eta h = \sigma$. Hence σ and η are conjugate.

Exercise 2. Given a positive integer n , a sequence of positive integers $n_1, n_2, \dots, n_r$ such that $n_1 \geq n_2 \geq \dots \geq n_r$ and $n = n_1 + n_2 + \dots + n_r$ is called a partition of n . Let $p(n)$ denote the number of partitions of n . Show that the number of conjugate classes in S_n is $p(n)$.

Note. Since $1=1$ is the only partition of 1, $p(1)=1$, $2=2$ and $2=1+1$ are two partitions of 2, so $p(2)=2$. $3=3$, $3=2+1$ and $3=1+1+1$ are three partitions of 3, so $p(3)=3$. We leave it to the readers to verify that $p(4)=5$, $p(5)=7$ and $p(6)=11$.

Solution. Let $\sigma \in S_n$. We can write σ as a product of pairwise disjoint cycles $(a_1 a_2 \dots a_{r_1}) (b_1 b_2 \dots b_{r_2}) \dots (c_1 c_2 \dots c_{r_k})$ such that $r_1 \geq r_2 \geq \dots \geq r_k$ and $r_1 + r_2 + \dots + r_k = n$. This is unique expression and so given any cycle we get a unique partition.

By Exercise 1, two permutations are conjugate if and only if they are similar, in other words they give rise to same partition. So corresponding to a conjugate class we get a unique partition of n . Conversely given a partition $n_1 \geq n_2 \geq \dots \geq n_s$, $n_1 + n_2 + \dots + n_s = n$ there exists a permutation f which has a cyclic decomposition of the type $(\alpha_1 \alpha_2 \dots \alpha_{n_1})(\beta_1 \beta_2 \dots \beta_{n_2}) \dots (\delta_1 \delta_2 \dots \delta_{n_s})$.

Each $g \in S_n$ similar to f is conjugate of f , and every permutation in S_n conjugate to f is similar to f . In this manner given any permutation we can associate a unique conjugate class viz. conjugate class of f .

Hence there is 1-1 correspondence between conjugate classes in S_n and partitions of n . As a consequence the number of conjugate classes in S_n is equal to $p(n)$.

Exercise 3. Let G be a finite group. Suppose an automorphism f of G sends more than three-quarters of the elements of G onto their inverses. Prove that G is Abelian and $f(x) = x^{-1} \forall x \in G$.

Solution. Suppose G is not Abelian then $Z = Z(G) \neq G$. This gives $o(Z) \leq \frac{1}{2} o(G) \Rightarrow o(G - Z) \geq \frac{o(G)}{2}$. So because of hypothesis there exists $x \in G - Z$ such that $f(x) = x^{-1}$.

Again $x \in G - Z \Rightarrow N(x) \neq G \Rightarrow o[N(x)] \leq \frac{o(G)}{2} \Rightarrow o[G - N(x)] \geq \frac{o(G)}{2}$.

As f sends more than three-quarters of the elements of G onto their inverses, there exists $y \in G - N(x)$ such that $f(y) = y^{-1}$.

Put $S = \{y \in G - N(x) \mid f(y) = y^{-1}\}$. Given condition on f implies that $o(S) > \frac{o(G)}{4}$. The set $xS = \{xs \mid s \in S\}$ has all elements distinct and so $o(xS) = o(S) > \frac{o(G)}{4}$.

We claim that for any $z \in xS$, $f(z) \neq z^{-1}$. Otherwise let $z \in xS$ be such that $f(z) = z^{-1} \Rightarrow f(xs) = (xs)^{-1}$ for some $s \in S \Rightarrow f(x)f(s) = s^{-1}x^{-1} \Rightarrow x^{-1}s^{-1} = s^{-1}x^{-1} \Rightarrow xs = sx \Rightarrow s \in N(x)$ which is against the definition of S . Thus our claim is established.

So there are more than $\frac{o(G)}{4}$ elements (namely, the elements of xS) which cannot be mapped onto their inverses. This is against the condition imposed on f , hence $G = Z$.

Finally let $H = \{x \in G \mid f(x) = x^{-1}\}$. H is a subgroup of G as G is Abelian. So if $H \neq G$, we shall get $o(H) \leq \frac{o(G)}{2}$, i.e. the number of elements of G which are mapped onto their inverses is less than or

equal to $\frac{o(G)}{2}$. This is against the hypothesis. Hence $H=G$. In other words $f(x)=x^{-1} \forall x \in G$.

Exercise 4. Show that for $n \geq 3$, $Z(S_n)=\{I\}$.

Solution. Suppose $\sigma \neq I \in Z(S_n)$.

We can write $\sigma = \sigma_1 \sigma_2 \dots \sigma_k$ where σ_i is a cycle of length $n_i (\geq 2)$ and these cycles are disjoint. Further $n_1 \geq n_2 \geq \dots \geq n_k$.

Case I. Let $\sigma_1 = (a_1 a_2 a_3 \dots a_n)$, $n \geq 3$.

Now $\sigma \in Z(S_n) \Rightarrow \sigma(a_1 a_2) = (a_1 a_2) \sigma$.

As each of $\sigma_2, \dots, \sigma_k$ is disjoint from σ_1 we get

$$\sigma_1(a_1 a_2) = (a_1 a_2) \sigma_1$$

$$\text{i.e. } (a_1 a_2 a_3 \dots a_n)(a_1 a_2) = (a_1 a_2)(a_1 a_2 a_3 \dots a_n).$$

This is absurd as L.H.S. sends a_n to a_1 , whereas R.H.S. sends a_n to a_2 .

Case II. Let $\sigma_1 = (a_1 a_2)$

Since $n \geq 3$ we can choose $a_3 \neq a_1, a_3 \neq a_2$.

Now $\sigma \in Z(S_n) \Rightarrow \sigma(a_1 a_2 a_3) = (a_1 a_2 a_3) \sigma$

$$\text{i.e. } (a_1 a_2) \sigma_2 \sigma_3 \dots \sigma_k (a_1 a_2 a_3) = (a_1 a_2 a_3) \sigma_2 \sigma_3 \dots \sigma_k.$$

As a_2 does not belong to any of cycles $\sigma_2, \sigma_3, \dots, \sigma_k$, L.H.S. sends a_1 to a_1 itself.

Because of same reasons RHS sends a_1 to a_3 .

This is again absurd. Hence our supposition that $\sigma \neq I \in Z(S_n)$ is wrong.

Consequently $Z(S_n) = \{I\}$.

PROBLEMS

- Show that the following pairs of permutations are similar in S_6 .
 - $(12), (13)$;
 - $(12), (34), (14), (23)$;
 - $(123), (45), (145), (23)$;
 - $(134), (256), (156), (342)$.
- Prove that the Klein 4-group is normal subgroup of A_4 .
- Prove that there exists no finite, non-trivial group every non-identity element of which commutes with exactly half the elements of the group.
- Let G be a group and $a \in G$ be such that a has only two conjugates. Using the fact that any subgroup of G of index 2 is normal, conclude that $N(a)$ is a normal subgroup of G .
- (a) Let H be an Abelian group of order p^n , p a prime number. Show by induction on n that there exist subgroups $H = H_0 > H_1 > H_2 > \dots > H_n = (e)$ of orders $p^n, p^{n-1}, \dots, p, p^0 = 1$ respectively.

(b) Using (a) show that if $o(G) = p^n$ (G may not be Abelian) there exist subgroups N_i ($i=0, 1, 2, \dots, n$) of orders p^i ($i=0, 1, 2, \dots, n$) such that $N_0 > N_1 > \dots > N_n = \{e\}$ and each N_{i+1} is normal in N_i for $i=0, 1, 2, \dots, n-1$. [Hint. $Z(G) \neq \{e\}$.]

6. If G is a non-Abelian group of order p^3 , where p is a prime number, show that $o[Z(G)] = p$.

[Hint. Use the fact that if $\frac{G}{Z(G)}$ is cyclic, then G is Abelian.]

7. If G is a group of order p^n , p a prime number and $H \neq G$ is a subgroup of G , show that there exists an $a \in G$, $a \notin H$ such that $aHa^{-1} = H$.

8. If G is a group of order p^n , p a prime number and if $N \neq \{e\}$ is a normal subgroup of G , prove that $N \cap Z(G) \neq \{e\}$ where $Z(G)$ is the centre of G .

9. Give an example of a group G of order p^n , p a prime number and a subgroup $H \neq \{e\}$ of G such that $H \cap Z(G) = \{e\}$ where $Z(G)$ is the centre of G .

10. Give an example of an automorphism of a non-Abelian finite group G which sends exactly three-quarters of the elements of G to their inverses.

11. Let N be a subgroup of G , $N \subseteq Z(G)$. The mapping $x \rightarrow x^m$ is 1-1 on N and 1-1 on G/N . Show that it is 1-1 on G . Further show that if the mapping $x \rightarrow x^m$ is onto on N and onto on G/N then it is onto on G .

REMARK. By a mapping f on a set A we mean f from A to A .

12. In any group G , the subset F of all elements which have only a finite number of conjugates in G , is a characteristic subgroup. Prove!

13. A group G is called an *FC group* (finite conjugate group) if and only if each $x \in G$ has only a finite number of conjugates. Prove the following:

(i) A finite group is an *FC group*.

(ii) An Abelian group is an *FC group*.

(iii) If $[G : Z(G)]$ is finite then G is an *FC group*.

(iv) A subgroup or a quotient group of an *FC group* is an *FC group*.

14. Show that $Z(G) = \cap \{C(x) \mid x \in G\}$, where $C(x) = \{y \in G \mid yx = xy\}$.

15. Let x, y be elements of a group G , then prove

(i) $C(x) \subseteq C(y)$ if and only if $y \in Z[C(x)]$.

(ii) $C(x) \subseteq C(y)$ if and only if $Z[C(x)] \supseteq Z[C(y)]$.

16. Determine all the conjugate classes in S_4 .

17. Let G be a group containing an element of order $n > 1$ and

exactly two conjugacy classes. Prove that C is a cyclic group of order 2.

18. Let G be a finite group with subgroup H of index 2. If $x \in H$ has m conjugates in G then show that x has either m or $\frac{m}{2}$ conjugates in H .

[Hint. Use Problem 25 after Theorem 2.42, that if A, B are subgroups of a finite group G such that $G=AB$ then $(G:B)=(A:A \cap B)$].

19. Prove that any element σ in S_n which commutes with $(1, 2, \dots, r)$ is of the form $(1, 2, \dots, r)^i T$ where $i=0, 1, 2, 3, \dots, r-1$ and T is a permutation leaving each of $1, 2, \dots, r$ fixed.

20. (a). Find the number of conjugates of $(1, 2)(3, 4)$ in S_n , $n \geq 4$.

(b) Find the form of all elements commuting with $(1, 2)(3, 4)$ in S_n .

Ans. 20 (a) $\frac{n(n-1)(n-2)(n-3)}{8}$ (b) $(12)^i (34)^j T$ and $(12)^i (34)^j (13)(24) T$, where $i=0, 1; j=0, 1$ and T leaves 1, 2, 3 and 4 fixed.

21. Show that $(1\ 2\ 3\ 4\ 5)$, $(1\ 2\ 3\ 5\ 4)$ are conjugate in S_5 but not in A_5 .

22. Let H be a proper subgroup of a finite group G show that G contains at least one element which is not in any conjugate of H .

STRUCTURE THEORY OF GROUPS

In this chapter we discuss some miscellaneous concepts in group theory, each of which has proved to be of great significance in the understanding of the inner structure of groups.

1. Direct Products

Let G be a group. We have already discussed the subgroups and quotient groups of G (Chapters 2 & 3). In a sense these are new groups constructed from G . There are many other methods of constructing new groups from one or more given groups, and these groups are different from the class of subgroups and quotient groups. One of the most important method is the method of formation of direct products of groups. In this section we discuss the concept of direct product and give some of its applications.

Theorem 4.1. Let $G_1, G_2, \dots, G_n$ be n groups and $G = G_1 \times G_2 \times \dots \times G_n$, their cartesian product. Let $a = (a_1, a_2, \dots, a_n)$, $b = (b_1, b_2, \dots, b_n) \in G$: define $ab = (a_1 b_1, a_2 b_2, \dots, a_n b_n)$. Then G is a group under this composition.

Proof: Let $a = (a_1, a_2, \dots, a_n)$, $b = (b_1, b_2, \dots, b_n)$, $c = (c_1, c_2, \dots, c_n)$, $\in G$.

For each $i = 1, 2, \dots, n$, let e_i be the identity of G_i , and $e = (e_1, e_2, \dots, e_n)$.

(1) *Closure.* By definition $ab = (a_1 b_1, a_2 b_2, \dots, a_n b_n)$. Since $a_i b_i \in G_i$, $1 \leq i \leq n$, we get $ab \in G$. This proves that G is closed under the composition.

(2) *Associativity.* Now $a(bc) = (a_1 a_2, \dots, a_n) (b_1 c_1, b_2 c_2, \dots, b_n c_n) = [a_1(b_1 c_1), a_2(b_2 c_2), \dots, a_n(b_n c_n)]$.

Similarly $(ab)c = [(a_1 b_1)c_1, (a_2 b_2)c_2, \dots, (a_n b_n)c_n]$.

Since $a_i(b_i c_i) = (a_i b_i)c_i \forall i$. We get $(ab)c = a(bc)$.

(3) *Existence of Identity.* Since $ae = (a_1 e_1, a_2 e_2, \dots, a_n e_n) = (a_1, a_2, \dots, a_n) = (e_1 a_1, e_2 a_2, \dots, e_n a_n) = ea$, we get $ae = a = ea \forall a \in G$. Hence e is the identity of G .

(4) *Existence of Inverses.* For given $a=(a_1, a_2, \dots, a_n) \in G$. Let $a'=(a_1^{-1}, a_2^{-1}, \dots, a_n^{-1})$. Then $aa'=(a_1a_1^{-1}, a_2a_2^{-1}, \dots, a_na_n^{-1})=(e_1, e_2, \dots, e_n)=e$. Similarly $a'a=e$. Hence $a^{-1}=a' \in G$.

This proves that G is a group. ■

Having proved the above theorem we now define external direct product of groups.

Definition 4.2. (External direct product). Let $G_1, \dots, G_n$ be a finite number of groups and $G=G_1 \times G_2 \times \dots \times G_n$. Then G is a group under the binary composition defined by $ab=(a_1b_1, a_2b_2, \dots, a_nb_n)$ for all $a=(a_1, a_2, \dots, a_n), b=(b_1, b_2, \dots, b_n)$ in G .

This group is called the *External direct product* of groups $G_1, G_2, \dots, G_n$.

If all the groups are Abelian and their compositions are all denoted additively, then usually, we also denote the composition in G additively and refer to G as a *direct sum* of $G_1, G_2, \dots, G_n$. Further we write $G=G_1 \oplus G_2 \oplus \dots \oplus G_n$.

EXAMPLE 1. Let $G=\mathbf{Z} \times \mathbf{Z}$, where $\mathbf{Z}$ is the additive group of integers. For $x=(n_1, m_1), y=(n_2, m_2) \in G$, $x+y=(n_1+n_2, m_1+m_2)$. G is a group under the addition composition. Here $(0, 0)$ is the identity and for any $(n, m) \in G$, $(-n, -m)$ is the inverse.

EXAMPLE 2. Let $\mathbf{Z}$ be the additive group of integers, and C^* , the multiplicative group of non-zero complex numbers.

Let $G=\mathbf{Z} \times C^*$. For $x=(x_1, z_1), y=(x_2, z_2) \in G$; define $xy=(x_1+x_2, z_1z_2)$. G is a group under this binary composition. $(0, 1)$ is the identity of G and for each $(n, x) \in G$, $(-n, x^{-1})$ is its inverse.

EXAMPLE 3. Let $G_1=\langle a \rangle$ and $G_2=\langle b \rangle$ be two cyclic groups of order 2 and 3 respectively. If e_1, e_2 are the identities of G_1 and G_2 respectively, then $G_1=\{e_1, a\}$ and $G_2=\{e_2, b, b^2\}$. Further $G=G_1 \times G_2=\{(e_1, e_2), (e_1, b), (e_1, b^2), (a, e_2), (a, b), (a, b^2)\}$.

For any $x=(a_1, a_2), y=(b_1, b_2) \in G$, define $xy=(a_1b_1, a_2b_2)$. Under this composition G becomes a group. Take $u=(a, b)$, we see that $u^2=(a^2, b^2)=(e_1, b^2), u^3=(a^3, b^3)=(a, e_2), u^4=(a^4, b^4)=(e_1, b), u^5=(a^5, b^5)=(a, b^2), u^6=(a^6, b^6)=(e_1, e_2)$. Hence $G=\langle u \rangle$ is a cyclic group of order 6.

Theorem 4.3. Let $G_1, G_2, \dots, G_n$ be a finite number of groups and $G=G_1 \times G_2 \times \dots \times G_n$, their direct product. For each $i(1 \leq i \leq n)$, let e_i be the identity of G_i . Then

(1) for each $i, G_i'=\{(e_1, e_2, \dots, e_{i-1}, a_i, e_{i+1}, \dots, e_n) \mid a_i \in G_i\}$ is a normal subgroup of G isomorphic to G_i .

(2) for every $i \neq j$, given $a_i' \in G_i', a_j' \in G_j', a_i' a_j' = a_j' a_i'$.

(3) each $a \in G$ is expressible uniquely as $a_1' a_2' \dots a_n'$, where $a_i' \in G_i'$, in the sense that if $a_1' a_2' \dots a_n' = b_1' b_2' \dots b_n'$ with $a_i', b_i' \in G_i'$, then

$$a_i' = b_i' \quad \forall i.$$

Proof: For each i , define $f_i : G_i \rightarrow G$ by $f_i(a_i) = (e_1, e_2, \dots, e_{i-1}, a_i, e_{i+1}, \dots, e_n) \quad \forall a_i \in G_i$.

For any $a_i, b_i \in G_i$, $f_i(a_i b_i) = (e_1, e_2, \dots, e_{i-1}, a_i b_i, e_{i+1}, \dots, e_n) = (e_1, e_2, \dots, e_{i-1}, a_i, e_{i+1}, \dots, e_n) (e_1, e_2, \dots, e_{i-1}, b_i, e_{i+1}, \dots, e_n) = f_i(a_i) f_i(b_i)$. Thus f_i is a homomorphism.

$$\begin{aligned} \text{Further } a_i \in \text{Ker } f_i &\Leftrightarrow f_i(a_i) = e = (e_1, e_2, \dots, e_i, \dots, e_n) \\ &\Leftrightarrow (e_1, e_2, \dots, e_{i-1}, a_i, e_{i+1}, \dots, e_n) \\ &= (e_i, e_2, \dots, e_{i-1}, e_i, e_{i+1}, \dots, e_n) \\ &\Leftrightarrow a_i = e_i. \end{aligned}$$

Hence $\text{Ker } f_i = (e_i)$ and f_i is a monomorphism of G_i into G . Thus $f_i(G_i) \cong G_i$. From the definition of G_i' we see that $f_i(G_i) = G_i'$. Consequently $G_i' \cong G_i$. Consider any $x = (x_1, x_2, \dots, x_n) \in G$, $a_i' \in G_i'$. Then $x^{-1} a_i' x = (x_1^{-1}, x_2^{-1}, \dots, x_n^{-1}) (e_1, e_2, \dots, e_{i-1}, a_i, e_{i+1}, \dots, e_n) (x_1, x_2, \dots, x_n) = (e_1, e_2, \dots, e_{i-1}, x_i^{-1} a_i x_i, e_{i+1}, \dots, e_n) \in G_i'$. Hence G_i' is a normal subgroup of G .

(2) Consider $a_i' = (e_1, e_2, \dots, e_{i-1}, a_i, e_{i+1}, \dots, e_n) \in G_i'$,
 $a_j' = (e_1, e_2, \dots, e_{j-1}, a_j, e_{j+1}, \dots, e_n) \in G_j'$ with $i \neq j$.

Then $a_i' a_j' = (e_1^2, e_2^2, \dots, a_i e_j, e_i^2 e_{j+1}, \dots, e_{j-1}^2, e_j a_i, e_{j+1}^2, \dots, e_n^2) = (e_1, e_2, \dots, e_{i-1}, a_i, e_{i+1}, \dots, e_{j-1}, a_j, e_{j+1}, \dots, e_n)$.

Similarly $a_j' a_i' = (e_1, e_2, \dots, e_{i-1}, a_i, e_{i+1}, \dots, a_{j-1}, a_j, e_{j+1}, \dots, e_n)$.

Hence $a_i' a_j' = a_i' a_j'$.

(3) Consider $a = (a_1, a_2, \dots, a_n) \in G$.

Clearly $a = (a_1, e_2, \dots, e_n) (e_1, a_2, e_3, \dots, e_n) \dots (e_1, e_2, \dots, e_{n-1}, a_n) = a_1' a_2' \dots a_n'$, where $a_1' = (a_1, e_2, \dots, e_n) \in G_1'$, $\dots$, $a_n' = (e_1, e_2, \dots, e_{n-1}, a_n) \in G_n'$.

We show that for a given a ; $a_1', a_2', \dots, a_n'$ are unique. Let also $a = b_1' b_2' \dots b_n'$ with $b_i' \in G_i'$. By definition $b_1' = (b_1, e_2, \dots, e_n)$, $b_2' = (e_1, b_2, e_3, \dots, e_n)$, $\dots$, $b_n' = (e_1, e_2, \dots, e_{n-1}, b_n)$ for some $b_1 \in G_1, b_2 \in G_2, \dots, b_n \in G_n$. This gives that $a = (b_1, b_2, \dots, b_n)$. As given, $a = (a_1, a_2, \dots, a_n)$. Consequently $(a_1, a_2, \dots, a_n) = (b_1, b_2, \dots, b_n) \Rightarrow a_i = b_i \quad \forall i \Rightarrow a_i' = b_i'$.

This completes the theorem. ■

Let us try to look upon the above theorem from a different angle, we started with groups $G_1, G_2, \dots, G_n$ and constructed a group G with the following properties:

(i) G has subgroups $G_1', G_2', \dots, G_n'$.

(ii) For every $i \neq j$, if $a_i' \in G_i', a_j' \in G_j'$, then $a_i' a_j' = a_j' a_i'$.

(iii) Each element $a \in G$ is expressible uniquely as $a = a_1' a_2' \dots a_n'$, $a_i' \in G_i' \quad \forall 1 \leq j \leq n$.

Now let G be any group (don't confuse G with the above direct

Proof: By (2), $H_1 \cap \prod_{j=2}^n H_j = (e) \Rightarrow H_1 \cap K = (e)$.

Since K is a product of normal subgroups, K is itself normal. Further $H_1 K = H_1 H_2 \dots H_n = G$. Hence G is the internal direct product of H_1 and K . ■

REMARK. As seen before if a group G is the internal direct product of its subgroups $H_1, H_2, \dots, H_n$, then $G \cong H_1 \times H_2 \times H_3 \times \dots \times H_n$. Again if a group G is an external direct product of groups $G_1, G_2, \dots, G_n$ then G contains subgroups $G'_1, \dots, G'_n$ such that $G'_i \cong G_i$ for all i and G is the internal direct product of $G_1, G_2, \dots, G_n$. As is customary for us to regard two isomorphic groups to be essentially equal, we can regard internal direct product, same as external direct product and call it simply a *direct product*.

Thus if we say that G is a direct product of groups $G_1, G_2, \dots, G_n$ then each G_i is a normal subgroup of G , $G = \prod_{i=1}^n G_i$ and $G_i \cap \prod_{j \neq i} G_j = (e) \forall i=1, 2, \dots, n$.

Suppose now G is direct product of groups $G_1, G_2, \dots, G_n$, for each $i=1, 2, \dots, n$, G_i is called a *direct factor* of G . In case G is direct sum of groups $G_1, G_2, \dots, G_n$; then, for each $i=1, 2, \dots, n$, G_i is called a *direct summand* of G .

Now we give some examples of internal direct products (sums).

EXAMPLE 4. Let $G = \langle a \rangle$ be a cyclic group of order 6. Let $H = \{e, a^2, a^4\}$, $K = \{e, a^3\}$. Using the fact that $o(a) = 6$, one can immediately conclude that H and K are 'subgroups' of G of orders 3 and 2 respectively. Observe (i) $H \cap K = \{e\}$, (ii) As G is Abelian H and K are both normal subgroups and (iii) $HK = \{ee, ea^3, a^2e, a^2a^3, a^4e, a^4a^3\} = \{e, a^3, a^2, a^5, a^4, a\} = G$, since $a^6 = e$.

Hence G is the internal direct product of H and K (In Theorem 4.6 take $n=2$, $H_1=H$, $H_2=K$).

EXAMPLE 5. In chapter 2.1, Problem 6 (iv), Klein four group V was defined. It consists of four elements $1, a, b, c=ab$, such that $a^2=b^2=(ab)^2=1$. This group is Abelian. $H=\{1, a\}$, $K=\{1, b\}$ are subgroups of V . These are normal, $H \cap K = \{1\}$ and $G=HK$. Hence G is a direct product of H and K .

Let G be a finite additive Abelian group. Let $o(G) = n = p_1^{\alpha_1} p_2^{\alpha_2} \dots p_t^{\alpha_t}$ be the decomposition of n , the order of G , into distinct prime powers. For each prime number $p \mid o(G)$, let $S(p) = \{x \in G \mid p^k x = 0, \text{ for some } k > 0\}$. As shown in Lemma 4.7, $S(p)$ is a subgroup of G . This subgroup $S(p)$ is called a *Sylow p -subgroup* of G . We shall show that G is the (internal) direct sum of its Sylow subgroups.

Consequently the structure of G will be known completely if we know the structure of its Sylow subgroups. The problem of determination of structure of Sylow subgroups finds its proper place in the next section.

Lemma 4.7. For any prime $p \mid o(G)$, where G is an (additive) Abelian group, the set $S(p)$ is a subgroup of G .

Proof: Let 0 be the additive identity of G . As $p0=0$, $0 \in S(p)$. Consequently $S(p) \neq \emptyset$. Now $x, y \in S(p) \Rightarrow p^k x = 0$. $p^l y = 0$ for some $k, l > 0$; Let $n = \max(k, l)$ then $p^n x = p^{n-k} p^k x = 0$ and similarly $p^n y = 0$. Thus $p^n(x-y) = p^n x - p^n y = 0 \Rightarrow x-y \in S(p)$. Hence $S(p)$ is a subgroup of G . ■

Lemma 4.8. Let x be any element of order mn in a group G such that $(m, n) = 1$. Then there exist $y, z \in G$ such that $x = yz = zy$, $o(y) = m$ and $o(z) = n$.

Proof: As $(m, n) = 1$, there exist integers t, s such that $1 = tm + sn$. Let $y = x^{sn}$, $z = x^{tm}$. Since y and z are both powers of same element, namely x , $yz = zy = x^{sn} x^{tm} = x^{tm+sn} = x^1 = x$.

Now $y^m = (x^{sn})^m = (x^{nm})^s = e$, since $o(x) = mn$. This gives that $o(y) = m_1 \mid m$. Similarly $z^n = e$ and $o(z) = n_1 \mid n$.

Then $(yz)^{m_1 n_1} = y^{m_1 n_1} z^{n_1 m_1} = ee = e \Rightarrow o(x) \mid m_1 n_1 \Rightarrow mn \mid m_1 n_1$. However $m_1 \mid m$, $n_1 \mid n$, so we must have $m = m_1$, $n = n_1$. This proves that $o(y) = m$, $o(z) = n$. ■

Lemma 4.9. Let x be an element of a group G such that $o(x) = n = n_1 n_2 \dots n_r$, where $n_1, n_2, \dots, n_r$ are pairwise co-prime integers. Then $x = x_1 x_2 \dots x_r$ for some element $x_i \in G$, ($1 \leq i \leq r$) such that $o(x_i) = n_i$, $\forall i$, and each x_i is a power of x .

Proof: For $r=1$, we can take $x_1 = x$, and the result follows. We apply induction, suppose $r > 1$ and the result holds for $r-1$.

Let $n' = n_1 n_2 \dots n_{r-1}$, $m = n_r$. By hypothesis $(n_i, n_r) = 1 \forall i \neq r \Rightarrow (n', n_r) = 1 \Rightarrow (n', m) = 1$. Then (by Lemma 4.8) $x = yz = zy$ for some $y, z \in G$ such that $o(y) = n'$, $o(z) = m = n_r$. As seen in the previous lemma, y and z can be both taken to be powers of x . Now $o(y) = n_1 n_2 \dots n_{r-1}$. So by induction hypothesis $y = y_1 y_2 \dots y_{r-1}$, for some $y_i \in G$ such that $o(y_i) = n_i$ and y_i is a power of y ($i=1, 2, \dots, r-1$). Since y itself is a power of x , we get each y_i is a power of x . Taking $x_i = y_i \forall i=1, 2, \dots, r-1$ and $x_r = z$, we get $x = x_1 x_2 \dots x_r$ such that $o(x_i) = n_i$ and each x_i is a power of x . This proves the lemma. ■

REMARK. If the binary composition of G is denoted additively, then in the above lemma, we should write $x = x_1 + x_2 + \dots + x_r$.

Theorem 4.10. Any finite additive Abelian group is internal direct sum of its Sylow subgroups.

Proof: Let G be a finite (additive) Abelian group, $o(G)=n=p_1^{\alpha_1} p_2^{\alpha_2} \dots p_t^{\alpha_t}$ where p_i are distinct prime numbers and α_i are integers >0 ($1 \leq i \leq t$). Then $S(p_1), \dots, S(p_t)$ are the only Sylow subgroups of G , since $p_1, \dots, p_t$ are the only prime numbers dividing the order of G .

(I) Now let $x \in G$, then $o(x) \mid o(G) = p_1^{\alpha_1} \dots p_t^{\alpha_t}$ gives $o(x) = p_1^{\beta_1} p_2^{\beta_2} \dots p_t^{\beta_t}$ for some β_i such that $0 \leq \beta_i \leq \alpha_i$. Now $p_1^{\beta_1}, \dots, p_t^{\beta_t}$ are pairwise co-prime, so $x = x_1 + x_2 + \dots + x_t$ (Lemma 4.9) such that each x_i is of order $p_i^{\beta_i}$. Then $p_i^{\beta_i} x_i = 0$ gives $x_i \in S(p_i) \forall i$. This shows that if $x \in G$, then $x = x_1 + x_2 + \dots + x_t$ for some $x_i \in S(p_i)$, $1 \leq i \leq t$. Consequently $G = S(p_1) + S(p_2) + \dots + S(p_t)$.

(II) Consider $S(p_i) \cap \sum_{j \neq i} S(p_j)$.

$x \in S(p_i) \cap \sum_{j \neq i} S(p_j) \Rightarrow x \in S(p_i)$ and $x = x_1 + x_2 + \dots + x_{i-1} + x_{i+1} + \dots + x_n$ for some $x_j \in S_j (j \neq i)$. Now by definition $x \in S(p_i) \Rightarrow p_i^{k_i} x = 0$ for some $k_i > 0 \Rightarrow o(x) \mid p_i^{k_i}$.

Again for each $j \neq i$, $x_j \in S(p_j) \Rightarrow p_j^{k_j} x_j = 0$

Let $m = p_1^{k_1} p_2^{k_2} \dots p_{i-1}^{k_{i-1}} p_{i+1}^{k_{i+1}} \dots p_n^{k_n}$.

Then $m x_j = 0 \forall j \neq i$ and $m x = m(x_1 + x_2 + \dots + x_{i-1} + x_{i+1} + \dots + x_n) = m x_1 + m x_2 + \dots + m x_{i-1} + m x_{i+1} + \dots + m x_n = 0$ yields that $o(x) \mid m$. Thus, $o(x) \mid m$ and $o(x) \mid p_i^{k_i}$. However, $p_i^{k_i}$ and m are co-prime. Consequently $o(x) = 1$, i.e. $x = 0$. Hence $S(p_i) \cap \sum_{j \neq i} S(p_j) = (0)$.

Further, trivially each $S(p_i)$ is a normal subgroup G . Hence G is the internal direct sum of $S(p_i) (i=1, 2, \dots, n)$. This completes the proof. ■

PROBLEMS

1. If G is a group and H, K are subgroups of G such that $G = H \times K$ prove that $H \cong G/K$ and $K \cong G/H$.

[Hint. For any $x \in G$, $x = hk$; $h \in H, k \in K$. Define $f: G \rightarrow H, g: G \rightarrow K$ by $f(x) = h, g(x) = k$. Show that f and g are homomorphisms onto H and K respectively with K and H as their respective kernels].

2. Let $G = \langle a \rangle$ be any cyclic group of order mn , where $(m, n) = 1$. Let H and K be its subgroups of orders m and n respectively. Show that $G = H \times K$.

3. Let $G_1 = \langle a \rangle, G_2 = \langle b \rangle$ be two cyclic groups of orders m and n , such that $(m, n) > 1$. Show that $G = G_1 \times G_2$ is an Abelian group of order mn , which is not cyclic.

4. Let G_1, G_2, G_3 be any three groups. Show the following:

(i) $o(G_1 \times G_2) = o(G_1) o(G_2)$, (where each G_i is finite)

(ii) $G_1 \times G_2 \cong G_2 \times G_1$.

(iii) $G_1 \times (G_2 \times G_3) \cong G_1 \times G_2 \times G_3 \cong (G_1 \times G_2) \times G_3$.

5. By using Lagrange's Theorem or otherwise show that any group G of order 4 is Abelian. Further show that either G is cyclic or G is isomorphic to a direct product of two cyclic groups each of order 2.

6. Let G be a finite group having at least three elements in which $a^2 = e \forall a \in G$. Show that G is internal direct product of a finite number of subgroups each of order 2; and $o(G) = 2^n$ for some $n \geq 2$.

7. Let G be a finite Abelian group, show that each of its Sylow subgroups is a characteristic subgroup of G .

8. Let G be a finite Abelian group and let $H_1, H_2, \dots, H_t$ be a finite number of subgroups of G such that $G = H_1 H_2 \dots H_t$ and $o(G)$

$$= \prod_{i=1}^t o(H_i), \text{ prove that } G = H_1 \times H_2 \times \dots \times H_t.$$

9. $G_1, G_2, \dots, G_n$ are n groups show that $Z(G_1 \times G_2 \times \dots \times G_n) = Z(G_1) \times Z(G_2) \times \dots \times Z(G_n)$.

10. Prove that the group $\mathbb{Q}$ of all rational numbers under addition can not be written as a direct sum of two non-trivial subgroups.

11. Let N be a normal subgroup of a group G . If $G = H \times K$ where H and K are subgroups of G then prove that either N is abelian or N intersects H or K non-trivially.

12. If M and N are normal subgroups of a group G then show that $G/M \cap N$ is isomorphic to a subgroup of the (external) direct product of $G/M \times G/N$.

[Hint. Define $\theta : G \rightarrow G/M \times G/N$ by $\theta(x) = (Mx, Nx)$].

13. A group G is said to be *decomposable* if and only if $G = H \times K$, where subgroups H and K are such that $H \neq \{e\}$, $K \neq \{e\}$, e being the identity of G .

Show that a cyclic group of order 6 is decomposable while S_3 is not decomposable.

14. Prove that the multiplicative group R of non-zero real numbers is decomposable.

[Hint. $R = R^+ \times \{1, -1\}$.]

15. A group is said to be *periodic* if every element of G is of finite order.

Show that if $G_1, G_2, \dots, G_n$ are periodic groups then $G_1 \times G_2 \times \dots \times G_n$ is also a periodic group.

16. Prove that if H is a direct factor of a group G , then every

normal subgroup K of H is a normal subgroup of G .

17. Let G_1 and G_2 be two groups. Prove that there exists a monomorphism from $\text{Aut } G_1 \times \text{Aut } G_2$ to $\text{Aut } (G_1 \times G_2)$. In case $\{o(G_1), o(G_2)\} = 1$ this monomorphism is isomorphism.

Give an example to show that in general this monomorphism need not be an isomorphism.

18. Let G be a finite group of order $n(>1)$ and $o(\text{Aut } G) \geq (n-1)!$ then prove that $G \cong Z_2$ or Z_3 or $Z_2 \times Z_2$.

[Hint. Use problem 5 above.]

19. G is a group, consider

$$H = \{(g, g) \mid g \in G\}$$

Show that H is a subgroup of $G \times G$. Further H is normal subgroup of $G \times G$ if and only if G is Abelian.

20. Let G be a group and α is a homomorphism from G to G such that $\alpha^2 = \alpha$. If $\alpha(G)$ is a normal subgroup of G , show that $G = \alpha(G) \times \text{Ker } \alpha$.

2. Sylow's Theorems

The reader must have noticed that how frequently Lagrange's Theorem for finite groups or its consequences have been used. Unluckily the converse of Lagrange's theorem does not hold in general, as we know that the alternating group A_4 is of order 12, but A_4 does not contain any subgroup of order 6. One of the main purposes of this section is to establish some theorems giving sufficient conditions on a divisor k of the order of finite group G , such that G contains a subgroup of order k . Theorem 4.12 shows that the converse of Lagrange's theorem does hold for finite Abelian groups. Also we establish some theorems giving relations between some subgroups of special types, of a finite group. The most important theorems in this connection are due to the Norwegian Mathematician Sylow. First of all we prove a theorem due to Cauchy.

Theorem 4.11. (Cauchy's Theorem for Abelian groups). Suppose G is a finite Abelian group and $p \mid o(G)$, where p is a prime number. Then there exists $a(a \neq e) \in G$ such that $a^p = e$.

Proof: As $p \mid o(G)$, $o(G) = n_1 p$ for some $n_1 \geq 1$, we proceed by induction on n_1 . Let $n_1 = 1$, then $o(G) = p$. As every group of prime order is cyclic, G is a cyclic group i.e. $\exists a \in G$ such that $G = \langle a \rangle$ then $o(a) = o(G) = p$ (Theorem 2.33). Consequently $a \neq e$ and $a^p = e$. Hence the result holds for $n_1 = 1$.

To apply induction, let $n_1 > 1$ and suppose that the result is true for every Abelian group G' for which $o(G') = mp$ which $m < n_1$ i.e. suppose the result to be true for all Abelian groups G' such that $p \mid o(G')$

and $o(G') < o(G)$. Since $o(G) = n_1 p$ is not prime, G must contain a proper subgroup (Theorem 3.47(a)). Suppose that G has a proper subgroup N such that $p \mid o(N)$. As $o(N) < o(G)$, by induction hypothesis $\exists a \in N$ such that $a \neq e$ and $a^p = e$. This a is also in G . So the result holds if G has a proper subgroup N with $p \mid o(N)$. Suppose now that p does not divide order of any proper subgroup of G . Let H be any proper subgroup of G . Then $o(G) = o(G/H) o(H)$, gives $p \mid o(G/H)$, since $p \nmid o(H)$. Further $o(G/H) = o(G)/o(H) < o(G)$. By the induction hypothesis G/H contains an element bH of order p . Thus $bH \neq H$ and $H = (bH)^p = b^p H$. This gives $b^p \in H$. This further yields $(b^p)^{o(H)} = e$ (Corollary 2.35). Let $a = b^{o(H)}$. Then $a^p = e$. If $a = e$, then we get $o(bH) \mid o(H) \Rightarrow p \mid o(H)$. This is a contradiction. Hence $a (\neq e) \in G$ is such that $a^p = e$. ■

Theorem 4.12. *If G is a finite Abelian group and a positive integer k divides $o(G)$, then G contains a subgroup of order k .*

Proof: We establish this theorem by induction on $o(G)$. If $o(G) = 1$, then $k \mid o(G) \Rightarrow k = 1$. In this case G is its own subgroup of order k . To apply induction we suppose that $o(G) > 1$ and the result holds for all Abelian group of order $< o(G)$. We can suppose that $k > 1$, since for $k = 1$, (e) is the subgroup of G of order 1. Let p be a prime number such that $p \mid k$. Then $p \mid o(G)$; so by Cauchy's theorem, given above, $\exists a \in G$ such that $o(a) = p$. Let $H = \langle a \rangle$ be the cyclic subgroup generated by a ; then $o(H) = p$. Now G/H is an Abelian group such that $o\left(\frac{G}{H}\right) = \frac{o(G)}{o(H)} = \frac{o(G)}{p} < o(G)$. As $p \mid k$, we have $k = k_1 p$ and $k_1 \mid o(G/H)$. So by induction hypothesis G/H has a subgroup, say T , of order k_1 . However every subgroup of G/H is of the form K/H , where K is a subgroup of G containing H . In particular $T = K/H$, where K is a subgroup of G containing H , then $o(K) = o(K/H) o(H) = k_1 p = k$. This completes the proof. ■

Theorem 4.13. (Sylow's First Theorem). *Let G be a finite group of order $n = p^k q$ ($k \geq 1$) where p is a prime number and q is a positive integer such that $(p, q) = 1$. Then for each i , ($1 \leq i \leq k$), G has at least one subgroup of order p^i .*

Proof: Let $o(G) = p^k q$. If $o(G) = p$, then G is cyclic group of order p , and the theorem holds trivially. We now proceed by induction on $o(G)$. Suppose that the result holds for all groups T of order $< o(G)$ such that $p \mid o(T)$. Let $Z(G)$ be the centre of G . Following cases arise.

Case I. $p \mid o(Z(G))$. In this case by Cauchy's Theorem for Abelian groups, $\exists a \in Z(G)$ of order p . Since a commutes with all elements of G , $H = \langle a \rangle$ is a normal subgroup of G of order p . If $k = 1$, then

H is the desired subgroup of G of order p . Let $k > 1$, then $o(G/H) = p^{k-1}q$. By induction hypothesis $\bar{G}$ contains subgroups $\bar{H}_i$ of orders p^i ($i=1, 2, 3, \dots, k-1$). Now for each i , $\bar{H}_i = H_i/H$ for some subgroup H_i of G containing H . Then $o(H_i) = o(\bar{H}_i) \cdot o(H) = p^{i+1}$ ($i=1, 2, 3, \dots, k-1$). Writing $K_1, K_2, K_3, \dots, K_k$ for $H, H_1, H_2, \dots, H_{k-1}$ we see that $o(K_i) = p^i \forall i=1, 2, \dots, k$.

Case II. $p \nmid o[Z(G)]$.

$$\text{We have } o(G) = o[Z(G)] + \sum_{N(a) \neq G} \frac{o(G)}{o[N(a)]}$$

where the summation runs over element a taken one each from each conjugate class containing more than one element. Since $p \mid o(G)$ therefore if $p \mid \frac{o(G)}{o[N(a)]}$ for all a for which $N(a) \neq G$ we will get from the above equality; $p \mid o[Z(G)]$, which is against hypothesis of this case. Hence there exists an $a \in G$ such that $N(a) \neq G$ and $p \nmid \frac{o(G)}{o[N(a)]}$.

Now $o(G) = \frac{o(G)}{o[N(a)]} \cdot o[N(a)]$ and $p \nmid \frac{o(G)}{o[N(a)]}$ then $p^k \mid o(G)$ implies that $p^k \mid o[N(a)]$. But $o[N(a)] < o(G)$ so by induction hypothesis $N(a)$ contains subgroups H_i 's of order p^i ($1 \leq i \leq k$), which are also subgroups of G . This establishes the theorem. ■

Corollary 4.14. (Cauchy's Theorem for finite groups). If a prime number p divides the order of a finite group G , then G contains at least one element of order p .

Proof: By Sylow's first theorem, G contains a subgroup H of order p . Then H is cyclic; so $\exists a \in G$ such that $H = \langle a \rangle$. Hence $o(a) = o(H) = p$. ■

Corollary 4.15. If $p^k \mid o(G)$ and $p^{k+1} \nmid o(G)$, then G has a subgroup of order p^k .

Proof: It is immediate from Theorem 4.13. ■

Definition 4.16. (p -subgroup). Let p be a prime number. Then a subgroup H of a group G is called a p -subgroup if the order of each element of H is a power of p .

In particular if each element of G is of order, a power of some fixed prime number p , then G is called a p -group.

Let G be a finite group. If $o(G) = p^k$, then, as by Lagrange's theorem, for each $a \in G$, $o(a) \mid p^k$, $o(a)$ is a power of p . Consequently G is a p -group. Conversely if G is a p -group then no prime $q \neq p$ can divide $o(G)$, since otherwise G will contain an element of order q , some prime number different from p (Cauchy's Theorem). Consequently if G is a p -group, then $o(G) = p^k$ is a power of p . Thus

G is a p -group if and only if $o(G)$ is a power of p .

Definition 4.17. (Sylow p -subgroup of a finite group). Let G be a finite group and p , a prime number, such that $p^k \mid o(G)$ and $p^{k+1} \nmid o(G)$. Then any subgroup of G of order p^k is called a Sylow p -subgroup of G .

The above definition implies that all Sylow p -subgroups of G are of same order viz., p^k . Since any p -subgroup is of order, a power of p , we see that no p -subgroup H of G can contain properly a Sylow p -subgroup of G .

EXAMPLE 6. Let $G = \langle a \rangle$ be a cyclic group of order 8. Since $8 = 2^3$, a power of prime number, G is a p -group.

EXAMPLE 7. Consider S_3 . $o(S_3) = 6 = 3 \cdot 2$. Thus $3 \mid o(S_3)$ but $3^2 \nmid o(S_3)$. So any Sylow 3-subgroup of S_3 is of order 3. Now $A_3 = \{I, (123), (132)\}$ is the only subgroup of S_3 of order 3. Hence A_3 is the only Sylow 3-subgroup of S_3 . Again the highest power of 2 which divides $o(S_3)$ is 2 itself. In this way any Sylow 2-subgroup of S_3 is of order 2. The only subgroups of S_3 of order 2 are $G_1 = \{I, (12)\}$, $H_2 = \{I, (13)\}$ and $H_3 = \{I, (23)\}$. Thus these three subgroups are the only Sylow 2-subgroups of S_3 .

EXAMPLE 8. Consider A_4 ; $o(A_4) = 12 = 2^2 \cdot 3$. Thus any Sylow 2-subgroup of A_4 is of order 4 and any Sylow 3-subgroup of A_4 is of order 3. Now $V_4 = \{I, (12)(34), (13)(24), (14)(23)\}$ is a subgroup of A_4 of order 4. V_4 is the only Sylow 2-subgroup (See Sylow's Third Theorem). $H_1 = \{I, (123), (132)\}$ is a Sylow 3-subgroup of A_4 . Indeed A_4 has 4 Sylow 3-subgroups. (Determine them?)

EXAMPLE 9. The group $G = \{\pm 1, \pm i, \pm j, \pm k\}$ of Quaternions is a non commutative group of order 8. As $8 = 2^3$, G is a 2-group.

Definition 4.18. Let S be any non-empty subset of a group G and $x \in G$. Then $x^{-1}Sx$ denotes the set $\{x^{-1}sx \mid s \in S\}$ and $x^{-1}Sx$ is called a conjugate of S determined by x .

Definition 4.19. Let S be any non-empty subset of a group G and K , a subgroup of G . Then normalizer of S in K is the set $N_K(S) = \{x \in K \mid x^{-1}Sx = S\}$.

We simply write $N(S)$ for $N_G(S)$ and call $N(S)$, the normalizer of S .

The concept of conjugacy of two elements of a group G and the concept of normalizer $N(a)$ of an element $a \in G$ were introduced in Chapter 3. We can prove that the relation of conjugacy on the family of all non-void subsets of G is an equivalence relation and $N(S)$ is a subgroup of G . The proofs are on the same lines as for the relation of conjugacy of elements and the normalizer $N(a)$ of an

element $a \in G$. Trivially for any subgroup K of G , $N_K(S) = N(S) \cap K$ and hence $N_K(S)$ is also a subgroup of G . We wish to give a word of warning that for a subset S of G , $x^{-1}Sx = S$ need not imply $xs = sx \forall s \in S$. This can be seen by noticing that if G is a non-Abelian group, there exists $g \in G$ such that $gx \neq xg$ for some $x \in G$ i.e., $x^{-1}gx \neq g$. However we know that $g^{-1}Gg = G$. Again as for $N(a)$, the number of conjugates of any subset S is equal to the index of $N(S)$ in G . Further let S be a subgroup of G ; then for each $s \in S$, $s^{-1}Ss = S$ gives $S \subseteq N(S)$. By definition $x^{-1}Sx = S \forall x \in N(S)$. So that S is a normal subgroup of $N(S)$. If S is normal in any subgroup H of G then $x^{-1}Sx = S \forall x \in H$ and hence $H \subseteq N(S)$. This shows that $N(S)$ is the largest subgroup of G , in which S is normal. Consider a subgroup K of G . If $x, y \in K$, then $x^{-1}Sx = y^{-1}Sy \Leftrightarrow yx^{-1}Sxy^{-1} = S \Leftrightarrow xy^{-1} \in N_K(S) \Leftrightarrow N_K(S)x = N_K(S)y$. Hence the number of conjugates of S determined by elements of K is equal to the index of $N_K(S)$ in K .

Lemma 4.20. *Let P be a Sylow p -subgroup of a finite group G . Then P contains all elements in $N(P)$, which are of order, a power of p .*

Proof: Let $x \in N(P)$ be such that $x^{p^k} = e$ for some $k \geq 0$. Since P is normal in $N(P)$, $N(P)/P$ is a group, and $xP \in N(P)/P$. Then $(xP)^{p^k} = x^{p^k}P = P$ gives $o(xP) \mid p^k \Rightarrow o(xP) = p^t$ for some $t \geq 0$. Let $\bar{K}$ be the subgroup of $N(P)/P$ generated by $\bar{x} = xP$. Then $o(\bar{x}) = o(\bar{K}) = p^t$. Now $\bar{K} = K/P$ for some subgroup K of $N(P)$ containing P . Let $o(P) = p^\alpha$. Then $o(K) = o(\bar{K})o(P) = p^{t+\alpha}$. Thus K is a p -subgroup of G containing P . Since no p -subgroup of G can contain a Sylow p -subgroup of G properly, we get $K = P$. Hence $xP = P$ and $x \in P$. This proves the Lemma. ■

Now we establish Sylow's second and third theorems.

Theorem 4.21. *Let G be a finite group and p , a prime number such that $p \mid o(G)$. Then*

(a) *All Sylow p -subgroups of G are conjugates of one another. (Sylow's Second Theorem).*

(b) *Number of Sylow p -subgroups is of the form $1 + mp$, where m is some non-negative integer such that $(1 + mp) \mid o(G)$. (Sylow's Third Theorem).*

Proof: (a) Let $o(G) = p^k q$ where $(p, q) = 1$ and $k \geq 1$. Let S be any Sylow p -subgroup of G and M , the set of all conjugates of S . We want to show that every Sylow p -subgroup P of G is a conjugate of S i.e., $P \in M$. Suppose on the contrary there exists a Sylow p -subgroup P of G which is not a conjugate of S . Let P' be any Sylow p -subgroup of G , different from P . Since $o(P') = o(P) = p^k$, we get $P \not\subseteq P'$. This yields $\exists x \in P$ such that $x \notin P'$. Then $x^{-1}P'x \neq P'$,

for otherwise $x^{-1}P'x = P' \Rightarrow x \in N(P')$; now x being a member of P , is of order, a power of p , then by lemma 4.20, $x \in P'$. This is a contradiction. Thus the number of conjugates of P' determined by elements of P is more than one. Since this number is equal to $[P : N_P(P')]$, $N_P(P') \neq P$. As $o(P) = p^k$, and $[P : N_P(P')] \mid p^k$ we get that the number of conjugates is p^l for some $l \geq 1$, hence it is a multiple of p .

On M define a relation ' $\sim$ ' as follows. $S_1 \sim S_2$ if and only if $S_1 = x^{-1}S_2x$ for some $x \in P$. This is an equivalence relation and it divides M into disjoint equivalence classes. Since for any $S' \in M$, $S' \neq P$, number of conjugates of S' determined by elements of P is a multiple of p (see previous paragraph), so the equivalence-class to which S' belongs has the number of members equal to a multiple of p . Since this holds for every equivalence class, we get that the number t of members of M is a multiple of p . However $t = [G : N(S)]$. Since $N(S) \supseteq S$ and $[G : S] = q$, we get $[G : N(S)] \mid q$. This gives $p \nmid t$, as $p \nmid q$. This contradicts the fact, proved above, that t is a multiple of p . Hence P must be a conjugate of S .

(b) Let M be the set of all Sylow p -subgroups of G . Take any $P \in M$. As in (a), define an equivalence relation ' $\sim$ ' on M as follows: $S_1 \sim S_2$ if and only if $S_1 = x^{-1}S_2x$ for some $x \in P$. Now if S is any Sylow p -subgroup of G different from P , as seen in (a), the number of conjugates of S determined by elements of P is a multiple of p .

This shows that the number of members in any equivalence class to which P does not belong is a multiple of p . Since $x^{-1}Px = P \forall x \in P$, P is the only member of the equivalence class to which P belongs. Hence the number of members of M is of the form $1 + mp$, where m is some non negative integer.

Since number of conjugates of a subgroup divides the order of group, $1 + mp$ divides the order of group.

This completes the proof. ■

REMARK. $(1 + mp) \mid p^k q$ and $(1 + mp, p^k) = 1 \Rightarrow ((1 + mp) \mid q$.

Theorem 4.22. Let G be a finite group and p , a prime number dividing the order of G . Then every p -subgroup B of G is contained in some Sylow p -subgroup of G .

Proof: Let M be the set of all Sylow p -subgroups of G . By Sylow's Third Theorem, number of members of M is equal to $1 + mp$ for some non-negative integer m . On M , define a relation ' $\sim$ ' as follows: For any $P_1, P_2 \in M$, $P_1 \sim P_2$ if and only if $P_1 = x^{-1}P_2x$ for some $x \in B$. This is an equivalence relation, and for any $P \in M$, the number of members of M in the conjugate class of P is same as the

number of conjugates of P determined by elements of B . If for some $P \in M, B \not\subseteq P$, then there exists $x \in B$ such that $x \notin P$. If $x^{-1}Px = P$, then as the order of x is a power of p (B is a p -subgroup) and $x \in N(P)$, we get $x \in P$ (Lemma 4.20.). This is a contradiction. So that the number t of conjugates of P is greater than 1. However $t = [B : N_B(P)]$ and $o(B) = p^l$ for some none-negative integer l , we get $t = p^s$ for some $s \geq 1$. Thus t is a multiple of p . Consequently if no P in M had contained B , then the number of members in each equivalence class would have been multiple of p . This would contradict the fact that the number of members of M is $1 + mp$. Hence there exists a Sylow p -subgroup P of G such that $B \subseteq P$. This proves the theorem. ■

APPENDIX TO SECTION 4.2

In this appendix we give alternative proofs of the theorems of Cauchy and Sylow.

Following proof of Cauchy's theorem is due to McKay.

Cauchy's Theorem

If a prime p divides the order of a finite group G then G has at least one element of order p .

Proof: Let S be the set of ordered p -tuples $(a_1, a_2, \dots, a_p)$ of elements of G , with the restriction that the product $a_1 a_2 \dots a_p = e$. If $n = o(G)$ then S has n^{p-1} elements (p -tuples), since the first $p-1$ of the a_i 's may be chosen arbitrarily and then the last a_p is the inverse of the product of previous a_i 's. Let σ be the permutation $(1\ 2\ 3\ \dots\ p)$ and consider it a transformation on the p -tuples. For instance $\{a_1, a_2, \dots, a_p\} \sigma = (a_2, a_3, \dots, a_p, a_1)$.

[Note. For convenience sake, mapping is written on right side.]

In other words σ permutes cyclically the elements of a p -tuple. We call two p -tuples α and β *equivalent* if $\alpha \sigma^i = \beta$ for some $i = 0, 1, \dots, p-1$. This equivalence is, in fact, an equivalence relation.

If $\alpha \in S$ the $\alpha \sigma \in S$; since $a_1 a_2 \dots a_p = e \Rightarrow \sigma_2 \dots a_p = a_1^{-1} \Rightarrow a_2 a_3 \dots a_p a_1 = e \Rightarrow a_3 a_4 \dots a_p a_1 = a_2^{-1} \Rightarrow a_3 a_4 \dots a_p a_1 a_2 = e$. Proceeding in this manner we see that $a_p a_{p-1} \dots a_2 a_1 = e$.

By means of this relation S may be partitioned into equivalence classes.

We claim that such a class of p -tuples either contain exactly one p -tuple or exactly p different p -tuples. Since σ generate a cyclic group H of order p , ($\sigma^p = I$, $\sigma^i \neq I$ for $0 < i < p$), no class can contain more than p different p -tuples. If a class contains less than p different p -tuples, then $\alpha \sigma^i = \alpha \sigma^j$ for α , a p -tuple of this class and i, j such

that $0 \leq i < j \leq p-1$. Then we can apply σ^{-i} or the right to get $\alpha = \tau \sigma^{j-i}$ i.e. $\alpha \sigma^k = \alpha$ where $k = j-i > 0$ and $1 \leq k \leq p-1$.

But $o(\sigma^k)$ must divide $o(H)$, hence $o(\sigma^k) = p$ and " σ^k generates H ." Thus every power of σ is a power of σ^k . Let $\sigma^i = (\sigma^k)^t$.

Then $\alpha \sigma^i = \alpha (\sigma^k)^t = \sigma^k \sigma^k \dots \sigma^k = \alpha$. Hence every power of σ takes α into itself i.e. all cyclic permutations of $(a_1, a_2, \dots, a_p)$ are equal. But that can happen only if all the a_i 's are equal; in which case the class contains only one p -tuple. This proves our claim.

Let r = no. of classes of S which contain only one p -tuple and s = no. of classes of S which contain p different p -tuples.

Then counting the number of elements of S in two different ways, we have

$$n^{p-1} = r + sp$$

Since by hypothesis $p \mid n$, p must divide r .

Each class of one p -tuple looks like $\{(a, a, \dots, a)\}$ since it is in S , $a^p = e$. Hence each class of one p -tuple gives a solution of $x^p = e$, and each solution gives such a class. Hence r is the number of solutions in G of $x^p = e$. But $p \mid r \Rightarrow r > 1$ so there are more than one solutions of $x^p = e$. One solution is obviously $x = e$. So there exists at least one element $a \in G$, $a \neq e$ such that $a^p = e$. Hence $o(a) = p$. ■

Double Cosets. Let H and K be two (not necessarily distinct) subgroups of a group G and let $x \in G$. The set HxK given by $\{h x k \mid h \in H, k \in K\}$ is called a *double coset*.

Lemma 1. Two double cosets HxK and HyK are either disjoint or identical.

Proof: Let $z \in HxK \cap HyK$.

Then $z = h x k = h' y k'$; $h, h' \in H$ and $k, k' \in K$.

This implies that $x = h^{-1} h' y k' k^{-1}$ and $y = (h')^{-1} h x k (k')^{-1}$

Now $t \in HxK \Rightarrow t = h_2 x k_2$; $h_2 \in H$ and $k_2 \in K$

$$= h_2 h^{-1} h' y k' k^{-1} k_2 \in HyK.$$

Conversely $u \in HyK \Rightarrow u = h_3 y k_3$; $h_3 \in H$ and $k_3 \in K$

$$= h_3 (h')^{-1} h x k (k')^{-1} k_3 \in HxK.$$

Hence $HxK = HyK$. ■

Let $a, b \in G$. $a \sim b$ if and only if $a = h b k$; $h \in H, k \in K$. It can be shown easily that $\sim$ is an equivalence relation on G .

Lemma 2. For subgroups H and K of a finite group G , $o(HxK)$

$$= \frac{o(H) o(K)}{o(x^{-1} H x \cap K)}.$$

Proof: Define $\theta : HxK \rightarrow x^{-1} HxK$ by $\theta(h x k) = x^{-1} h x k$ for all $h \in H, k \in K$. θ is evidently a 1-1 correspondence.

Hence $o(HxK) = o(x^{-1} HxK)$.

$$= \frac{o(x^{-1}Hx) o(K)}{o(x^{-1}Hx \cap K)}$$

But $o(x^{-1}Hx) = o(H)$.

So
$$o(HxK) = \frac{o(H) o(K)}{o(x^{-1}Hx \cap K)}. \blacksquare$$

Theorem (Frobenius). If G is a finite group, then

$$o(G) = \sum \frac{o(H) o(K)}{o(x^{-1}Hx \cap K)},$$

where summation on right hand side is taken over elements x chosen from disjoint double cosets HxK .

Proof: For each $a \in G$, $a \in HaK$.

This at once implies that $G = \bigcup_{x \in G} HxK$.

Writing disjoint union, we get

$$\begin{aligned} o(G) &= \sum o(HxK) \\ &= \sum \frac{o(H) o(K)}{o(x^{-1}Hx \cap K)}, \text{ by Lemma 2.} \end{aligned}$$

This proves the theorem. $\blacksquare$

Lemma 3. Let H and K be subgroups of a group G then HxK contains all right cosets of H in G of type Hxk , $k \in K$. Also HxK contains all left cosets of K in G of type hxK , $h \in H$. Further HxK consists of complete right cosets of H and complete left cosets of K .

Proof: Clearly any right coset of H of type Hxk , $k \in K$ is contained in HxK .

Let $Ha \subseteq HxK$ then $a \in Ha \Rightarrow a = h x k$; $h \in H$ and $k \in K$. This gives $Ha = H h x k = H x K$.

Thus the assertion about right cosets is established. The proof for left cosets is similar. $\blacksquare$

Before stating and proving next lemma, we make the following convention. Throughout this appendix, union is taken over disjoint subsets.

Lemma 4. G is a finite group. The number of distinct right cosets of H in HxK is $[K : x^{-1}Hx \cap K]$ and the number of distinct left cosets of K in HxK is $[x^{-1}Hx : x^{-1}Hx \cap K]$.

Proof: By lemma 3, every right coset of H in HxK is of type Hxk , $k \in K$.

Map θ taking Hxk to $x^{-1}Hxk$ is 1-1 from the family of all right cosets of H in HxK onto the family of all right cosets of $x^{-1}Hx$ in $x^{-1}HxK$.

Hence the number of distinct right cosets of H in HxK is equal to the number of distinct right cosets of $x^{-1}Hx$ in $x^{-1}HxK$ (1)

Let $A = x^{-1}Hx$ and $D = A \cap K$.

Now $K = \bigcup_{i=1}^r Du_i$, $u_i \in K$ and where $r = [K : D] = [K : x^{-1}Hx \cap K]$.

Consider $\{Au_i\}_{1 \leq i \leq r}$. These are right cosets of A in AK .

Let Ac be any right coset of A in AK . Then $c \in Ac \Rightarrow c \in AK \Rightarrow c = ak$, $a \in A$, $k \in K$.

So $Ac = Aak = Ak = Adu_s$ for some s , ($1 \leq s \leq r$ and $d \in D$) $= Au_s$, as $d \in D \Rightarrow d \in A$.

Also $Au_i = Au_j \Leftrightarrow u_i u_j^{-1} \in A$.

But $u_i u_j^{-1} \in K$ so $Au_i = Au_j \Leftrightarrow u_i u_j^{-1} \in A \cap K = D$.

Hence $Au_i = Au_j \Leftrightarrow i = j$.

Consequently the number of distinct right cosets of A in AK is equal to r .

Using (1) we get first part.

Proof for second part is similar. ■

Sylow's First Theorem

If G is a group of order $n = p^m q$ where p is a prime which does not divide q and m is integer ≥ 1 then G contains subgroups of order p^i ($i = 1, 2, \dots, m$) and each subgroup of order p^i ($i = 1, 2, \dots, m-1$) is a normal subgroup of at least one subgroup of order p^{i+1} .

Proof: By Cauchy's Theorem there is an element $a \in G$ of order p . Then $H = \langle a \rangle$ is a subgroup of order p . Suppose it has been proved that there exists a subgroup of order p^i , $1 \leq i \leq m-1$. Let P be a subgroup of order p^i .

Let $G = \bigcup_{i=1}^r Px_iP$, $x_1 = e$.

We can also write $G = \bigcup_{i=1}^s Py_i$, $y_i \in G$ and $y_1 = e$ where $s = \frac{o(G)}{o(P)} = p^{m-i}q$. As $i \leq m-1 \Rightarrow i < m$, $p \mid s$ (1)

Clearly $Px_i \subseteq Px_iP$ so we can decompose G into disjoint union of right cosets of P in Px_1P , disjoint union of right cosets of P in Px_2P , etc.

Let a_i be the number of distinct right cosets of P in Px_iP . Then $s = a_1 + a_2 + \dots + a_r$ (2)

Now by lemma 4, for all j , $1 \leq j \leq r$, $a_j = [P : P \cap x_j^{-1}Px_j]$. As $o(P) = p^i$, $P \cap x_j^{-1}Px_j \subseteq P \Rightarrow a_j = p^{m_j}$, $0 \leq m_j \leq i$. The only right coset of P in $Px_1P (= PeP = P)$ is P itself. So $a_1 = 1$.

If $m_j \geq 1$, $p \mid a_j$. In view of the fact that $p \mid s$, we get that the

number of a_j 's each equal to unity must be a multiple of p .

If $a_j=1$ then $x_j^{-1} P x_j = P \Rightarrow x_j \in N(P)$ and so $P x_j$ is a right coset of P in $N(P)$. Conversely if $P x_j$ is a right coset of P in $N(P)$, $x_j \in N(P)$. This in turn implies $x_j^{-1} P x_j = P$ so $a_j=1$.

Hence $[N(P) : P] = \text{number of } a_j\text{'s each equal to unity}$. Consequently $p \mid [N(P) : P]$.

Evidently P is normal subgroup of $N(P)$. By Cauchy's theorem $\frac{N(P)}{P}$ has an element of order p , which gives rise to a subgroup say $\frac{K}{P}$ of $\frac{N(P)}{P}$, of order p . By correspondence theorem K is subgroup of G containing P as normal subgroup. But $o(K) = \frac{o(K)}{o(P)} o(P) = p \cdot p^i = p^{i+1}$. This proves the theorem. ■

Corollary: Let G be a group of order $p^m s$ ($m > 1$) such that p is a prime and $p \nmid s$. Then every p -group which is a subgroup of G is contained in a Sylow p -subgroup of G .

Proof: Let H be a p -group which is subgroup of G . So $o(H) = p^i$, $i \leq m$. By repeated application of Sylow's first theorem we can find a sequence of subgroups $H_1, H_2, \dots, H_{m-i}$, of order $p^{i+1}, p^{i+2}, \dots, p^m$ respectively such that $H \subseteq H_1 \subseteq H_2 \subseteq \dots \subseteq H_{m-i}$. As $o(H_{m-i}) = p^m$, H_{m-i} is Sylow p -subgroup containing H . ■

Sylow's Second Theorem

All Sylow p -subgroups of a finite group G are conjugate.

Proof: Let $o(G) = p^m s$, $m > 1$ and $p \nmid s$.

Suppose there are two Sylow p -subgroups A and B of G which are not conjugate.

Let $G = \bigcup_{i=1}^r A x_i B$, $x_1 = e$. Then $o(G) = \sum_{i=1}^r o(A x_i B)$.

By Lemma 2, $o(A x_i B) = \frac{o(A) o(B)}{o(x_i^{-1} A x_i \cap B)} = p^{2m-i}$

where $o(x_i^{-1} A x_i \cap B) = p^i$.

Since B is not conjugate to A , $x_i^{-1} A x_i \cap B \subsetneq B \Rightarrow i < m$.

So $2m-i > m \Rightarrow p^{m+1} \mid o(A x_i B)$ for all $i=1, 2, \dots, r$.

Consequently $p^{m+1} \mid o(G)$. This is absurd as $p \nmid s$. Hence all Sylow p -subgroups are conjugate. ■

Sylow's Third Theorem

Number of Sylow p -subgroups of a finite group G is equal to $1+kp$ where k is an integer ≥ 0 such that $1+kp \mid o(G)$.

Proof: Let $o(G) = p^m s$, $m > 1$ and $p \nmid s$.

If P is a Sylow p -subgroup of G then $o(P) = p^m$.

$$\text{Now } G = \bigcup_{i=1}^k P x_i P \Rightarrow o(G) = \sum_{i=1}^k o(P x_i P) \quad \dots (1)$$

$$o(P x P) = \frac{[o(P)]^2}{o(x^{-1} P x \cap P)} \Rightarrow o(P x P) = \frac{p^{2m}}{o(x^{-1} P x \cap P)}.$$

Clearly $x^{-1} P x = P \Leftrightarrow x \in N(P)$. In other words $o(x^{-1} P x \cap P) = p^m$ if and only if $x \in N(P)$. Hence $p^{m+1} \mid o(P x P)$ if and only if $x \notin N(P)$.

Further $x \in N(P) \Leftrightarrow P x = x P \Leftrightarrow P x P = P x$. This in turn implies that in case $x \in N(P)$, $o(P x P) = o(P x) = o(P) = p^m$.

$$\begin{aligned} \text{By (1), we get } o(G) &= \sum_{x \in N(P)} o(P x P) + \sum_{x \notin N(P)} o(P x P) \\ &= \sum_{x \in N(P)} o(P x) + \sum_{x \notin N(P)} o(P x P) \\ &= o[N(P)] + \sum_{x \notin N(P)} o(P x P) \text{ as } N(P) = \bigcup_{x \in N(P)} P x. \end{aligned}$$

$$\text{Hence } o(G) = o[N(P)] + \sum_{x \notin N(P)} o(P x P) \quad \dots (2)$$

Each term in second summation is divisible by p^{m+1} so $o(G) = o[N(P)] + p^{m+1} u$, u is integer.

$$\text{This implies } \frac{o(G)}{o[N(P)]} = 1 + \frac{p^{m+1} u}{o[N(P)]}$$

As $N(P)$ is a subgroup of G , $\frac{p^{m+1} u}{o[N(P)]}$ is integer, say, t .

Thus $p^{m+1} u = t o[N(P)]$. It follows from (2) and the fact that $p^{m+1} \nmid o(G)$, $p^{m+1} \nmid o[N(P)]$. So $(p, t) \neq 1$, i.e. $p \mid t \Rightarrow t = kp$, k integer ≥ 0 .

$$\text{Therefore } \frac{o(G)}{o[N(P)]} = 1 + kp.$$

Now number of distinct conjugates of P in G is equal to $\frac{o(G)}{o[N(P)]}$.

So number of conjugates of P is $1+kp$.

By Sylow's Second Theorem all Sylow p -subgroups are conjugate.

Consequently the number of Sylow p -subgroups is $1+kp$. As $o[N(P)]$ is an integer, $1+kp \mid o(G)$. ■

WORKED-OUT EXERCISES

Exercise 1. If G is a finite group and p is a prime number dividing $o(G)$, prove that a Sylow p -subgroup P of G is a normal subgroup of G if and only if P is the only Sylow p -subgroup of G .

Solution. Suppose $o(G) = p^k q$ with $(p, q) = 1$. Then $o(P) = p^k$. Let $g \in G$; gPg^{-1} is a subgroup of G . As $gpg^{-1} = gp'g^{-1} \Rightarrow p = p', p, p' \in P$, we get that $o(gPg^{-1}) = o(P) = p^k$. By the hypothesis, this means that $gPg^{-1} = P$. Hence P is a normal subgroup of G .

Conversely, let P be a normal Sylow p -subgroup of G and Q , any Sylow p -subgroup of G . By Sylow's second theorem, Q and P must be conjugate, i.e., $Q = xPx^{-1}$ for some $x \in G$. But P is a normal subgroup of G , so $xPx^{-1} = P$. Hence $Q = P$.

Exercise 2. If G is a group of order $p^k s$, where p is a prime number such that $(p, s) = 1$, then show that each subgroup of order p^i , $i = 1, 2, \dots, k-1$ is a normal subgroup of at least one subgroup of order p^{i+1} .

(Note. This is considered, by many authors, to be a part of Sylow's first theorem).

Solution. We apply induction on $o(G)$. If $o(G) = p$ the result holds trivially.

Assume that the result holds for all groups T of order $< o(G)$ such that $p \mid o(T)$. Let $Z = Z(G)$ be the centre of G . Two cases arise:

Case I. $p \mid o(Z)$. By Cauchy's theorem for Abelian groups there exists $a \in Z$ such that $o(a) = p$. Then $H = \langle a \rangle$ is a normal subgroup of G of order p . If $k = 1$, p^i, p^{i+1} take up values 1 and p resp. and so our result holds. If $k > 1$, then $o(G/H) = p^{k-1}s < o(G)$, and by induction hypothesis every subgroup of G/H of order p^i is normal in at least one subgroup of G/H of order p^{i+1} for all $i = 1, 2, \dots, k-2$. Let K be a subgroup of G of order p^i , $i = 1, 2, \dots, k-1$. In case $H \subseteq K$, K/H is a subgroup of G/H of order p^{i-1} . By induction hypothesis there exists a subgroup L/H of G/H of order p^i such that K/H is normal in L/H . Then K is a normal subgroup of L and $o(L) = o(L/H)o(H) = p^i \cdot p = p^{i+1}$. Hence the result holds for all subgroups of G of order p^i , $i = 1, 2, \dots, k-1$. In case $H \not\subseteq K$, consider HK . Now $o(H \cap K) \mid o(H) \Rightarrow o(H \cap K) = 1$ or p . As $o(H \cap K) = p \Rightarrow H \cap K = H \Rightarrow H \subseteq K$ which is absurd, we get that $o(H \cap K) = 1$. As a consequence $o(HK) = o(H)o(K) = p^{i+1}$. Since $H \subseteq Z$, K is a normal subgroup of HK . (Prove!). Hence once again the result holds for all subgroups of G of order p^i , $i = 1, 2, \dots, k-1$.

Case II. $p \nmid o(Z)$. Let K be a subgroup of G of order p^i ($1 \leq i \leq k-1$). Now as $p \nmid o(Z)$, $o(K \cap Z) = 1$. This gives that $o(KZ)$

$=o(K)o(Z)=p^i$ $o(Z)<p^k$ $o(Z)\leq p^ks=o(G)$. By induction hypothesis there exists a subgroup N of KZ such that K is normal subgroup of N and $o(N)=p^{i+1}$ ($1\leq i\leq k-1$). But every subgroup of KZ is a subgroup of G . Hence the result holds for this case too.

Exercise 3. If a group G of order p^n , where p is a prime number and n is an integer ≥ 1 , has exactly one subgroup of order p^{n-1} , then prove that G is cyclic.

Solution. Let $o(G)=p^n$ where p is a prime number and n is a positive integer. Suppose that H is the subgroup of G of order p^{n-1} . Since $G\neq H$ there exist $a\in G$, $a\notin H$. Now $o(a)\mid p^n \Rightarrow o(a)=p^m$, $0\leq m\leq n$. If $m\neq n$, the cyclic subgroup T of G generated by a is of order p^m , $0\leq m\leq n-1$. By exercise 2, there exists a subgroup K of order p^{n-1} such that $T\subseteq K$. By the hypothesis $K=H$. So $\langle a\rangle=T\subseteq H$. But this implies that $a\in H$, a contradiction. Hence $m=n$ and $o(a)=p^n$. Consequently G is cyclic.

Exercise 4. Show that no group of order 30 is simple.

Solution. Let G be a group of order 30. Note that $30=2\cdot 3\cdot 5$. So G has at least a Sylow 2-subgroup, a Sylow 3-subgroup and a Sylow 5-subgroup. Number of Sylow 5-subgroups is $1+5k$ such that $(1+5k)\mid 6 \Rightarrow k=0$ or 1 . If $k=0$, then Sylow 5-subgroup is a normal subgroup of G . If $k=1$, number of Sylow 5-subgroups is six. Let $H_1, H_2, H_3, H_4, H_5, H_6$ be the Sylow 5-subgroups. Since $o(H_i\cap H_j)=1$ for $i\neq j$, these six Sylow 5-subgroups contain 24 elements of order 5 and there is one element of order 1. Thus only 5 remaining elements of G can be of order 2 or 3. Now number of Sylow 3-subgroups of G is $1+3m$ such that $(1+3m)\mid 10$. Then, $m=0$ or 3 . If $m=0$ we have only one Sylow 3-subgroup which must be normal. If $m=3$, we have 10 Sylow 3-subgroups. But these will contain 20 elements of order 3. This is absurd in view of the arguments given above. Hence G has either a normal subgroup of order 5 or a normal subgroup of order 3. Thus G is not simple.

Exercise 5. G is a group of order p^n , where p is a prime number and n is an integer ≥ 1 . Prove that any subgroup of order p^{n-1} is a normal subgroup of G .

Solution. Let H be a subgroup of G such that $o(H)=p^{n-1}$. By Exercise 2, there exists a subgroup K of G such that H is normal subgroup of K and $o(K)=p^n$. But this in turn implies $K=G$ and hence H is a normal subgroup of G .

REMARK. The result also follows by Exercise 4 after Theorem (2.42). But in the following we give a solution of this exercise, which does not depend on any of previous exercises.

We apply induction on n . For $n=1$, G is a cyclic group of prime

order and the only subgroup of order p^{n-1} i.e. 1, is $\{e\}$ where e is identity of G . This is trivially a normal subgroup of G . Suppose now that $n > 1$ and let the result be true for all subgroups T of order p^m , $1 \leq m < n$. Let H be a subgroup of G of order p^{n-1} . H is a normal subgroup of $N(H)$. If $N(H) \neq H$ then $o[N(H)] = p^n$. So $N(H) = G$ and as a consequence H is a normal subgroup of G .

Consider the case $N(H) = H$. As $Z = Z(G)$ is centre of G , $Z \subseteq N(H) \Rightarrow Z \subseteq H$. Since $p \mid o(G)$, there exists an element $a \in Z$ such that $o(a) = p$. The subgroup $K = \langle a \rangle$ is a normal subgroup of G , of order p . Now $o(H/K) = o(H)/o(K) = p^{n-1}/p = p^{n-2}$. By induction hypothesis H/K is a normal subgroup of G/K as $o(G/K) = p^{n-1}$. Then H is a normal subgroup of G . Hence the result follows.

Exercise 6. Show that no group of order 108 is simple.

Solution. Let G be a group of order 108. As $108 = 2^2 \cdot 3^3$, number of Sylow 3-subgroups is $1 + 3k$ such that $(1 + 3k) \mid 4$. Either $k = 0$ or $k = 1$. In case $k = 0$, there is only one Sylow 3-subgroup (which is of order 27). So it must be normal. Hence in this case G has a normal subgroup of order 27. Consequently G is not simple.

In case $k = 1$, there are four Sylow 3-subgroups of G , say H_1, H_2, H_3 and H_4 . As $o(H_1 \cap H_2) \mid o(H_1) \Rightarrow o(H_1 \cap H_2) = 1, 3, 9$. As $o(H_1 \cap H_2) = 1$ or 3 implies $o(H_1 H_2) = \frac{o(H_1)o(H_2)}{o(H_1 \cap H_2)} > 108$ which is absurd as $H_1 H_2 \subseteq G$, we get that $o(H_1 \cap H_2) = 9 = 3^2$.

By exercise 5, $H_1 \cap H_2$ is a normal subgroup of H_1 as well as of H_2 . Thus $H_1 \subseteq N(H_1 \cap H_2)$ and $H_2 \subseteq N(H_1 \cap H_2) \Rightarrow H_1 H_2 \subseteq N(H_1 \cap H_2) \Rightarrow o[N(H_1 \cap H_2)] \geq o(H_1 H_2) = \frac{o(H_1)o(H_2)}{o(H_1 \cap H_2)} = \frac{27 \cdot 27}{9} = 81$. As $N(H_1 \cap H_2)$ is a subgroup of G , $o[N(H_1 \cap H_2)] \mid 108$. This implies that $o[N(H_1 \cap H_2)] = 108 \Rightarrow N(H_1 \cap H_2) = G$. Now $H_1 \cap H_2$ is a normal subgroup of $N(H_1 \cap H_2)$. Hence $H_1 \cap H_2$ is a normal subgroup of G of order 9. As a consequence G is not a simple group.

PROBLEMS

- Let G be a finite group and p , a prime number dividing $o(G)$.
 - For any Sylow p -subgroup P , show that P is the only Sylow p -subgroup of $N(P)$.
 - Let P be a Sylow p -subgroup of G and H , any subgroup of G containing $N(P)$. Show that $N(H) = H$. In particular conclude that $N[N(P)] = N(P)$ [**Hint.** Apply Sylow's Second Theorem.]
- Prove that the following definition of Sylow p -subgroup is equivalent to the one given earlier.

Definition. Let G be a finite group and p be a prime number, then a subgroup P of G is called a *Sylow p -subgroup* of G if it satisfies the following conditions.

(1) P is p subgroup.

(2) No p -subgroup of G contains P properly.

[Hint. Use Theorem 4.22]

3. Show that if H is a normal subgroup of a finite group G and p , a prime dividing $o(G)$ such that $[G : H]$ is co-prime to p , then H contains all Sylow p -subgroups of G . Give an example to show that the result need not be true in case H is not normal.

[Hint. If p^k is the highest power of p dividing $o(G)$ then $p^k \mid o(H)$. Now use Sylow's Second Theorem.]

4. Show that a normal p -subgroup K of a finite group G is contained in every Sylow p -subgroup of G .

[Hint: Use Theorem 4.22 and Sylow's First Theorem.]

5. Prove that groups of order 15 and 35 are cyclic.

6. Give an example of a group of order p^3 , p a prime number, which is not commutative.

7. If $o(G)=36$, prove that G has 1 or 4 Sylow 3-subgroups.

8. If $o(G)=56$, prove that G has 1 or 8 Sylow 7-subgroups. In the latter case prove that G has a normal Sylow 2-subgroup.

9. Let G be an Abelian group and H, K , two finite subgroups of order m and n . Prove that G has a subgroup of order LCM of m and n .

[Hint. Since $m \mid o(HK)$ and $n \mid o(HK)$, if $k=LCM$ of m and n then $k \mid o(HK)$. Now use Theorem 4.12].

10. Prove that a group of order 48 must have a normal subgroup of order 8 or 16.

11. Prove that a group of order p^2q , where p and q are distinct primes has at least one normal Sylow subgroup.

[Hint. Use Sylow's Third Theorems.]

12. Let G be a finite group and p be a prime number dividing $o(G)$. If H is a normal subgroup of G , then prove that for any Sylow p -subgroup P of G , PH/H is a Sylow p subgroup of G/H . Conversely show that every Sylow p -subgroup of G/H is of the form PH/H where P is a Sylow p subgroup of G .

13. Show that a group G is a p -group if and only if there exists a normal subgroup N of G such that N and G/N both are p -groups.

14. Prove that if a group G of order 28 has a normal subgroup of order 4, then G must be Abelian.

15. Give a proof of Cauchy's theorem for finite groups (not necessarily abelian) which does not depend on Sylow's first theorem.

16. If each Sylow subgroup of G is normal then show that G is the direct product of its Sylow subgroups.

17. H is a subgroup of a finite group G and p is a prime number dividing the order of H . If P, Q are distinct Sylow p -subgroups of H and if S, T are Sylow p -subgroups of G such that $P \subseteq S$ and $Q \subseteq T$, show $S \neq T$.

18. If H is a subgroup of finite group G and p is a prime number dividing $o(H)$, show that number of Sylow p -subgroups of H is less than or equal to the number of Sylow p subgroups of G .

19. G is a finite group and p is a prime number dividing order of G . If A is a normal subgroup of G and H is a Sylow p -subgroup of G then prove that $A \cap H$ is a Sylow p -subgroup of A .

[Hint. Deny the result. There exists a Sylow p -subgroup P of A and a Sylow p -subgroup Q of G such that $A \cap H < P \subseteq Q$. By Sylow's second theorem, $H = xQx^{-1}$ for some $x \in G$. Then $xPx^{-1} \subseteq x(A \cap Q)x^{-1}$ as $P \subseteq A$. This will give $xPx^{-1} \subseteq (xAx^{-1}) \cap (xQx^{-1}) \Rightarrow xPx^{-1} \subseteq (xAx^{-1}) \cap H \Rightarrow xPx^{-1} \subseteq A \cap H$ as A is a normal subgroup of G . Hence $xPx^{-1} < P$. But xPx^{-1} is a Sylow p -subgroup of G and hence we get a contradiction.]

20. Give an example of a finite group G and a subgroup A of G , such that when p , a prime number, divides $o(G)$ and H is a Sylow p -subgroup of G , then $A \cap H$ is not a Sylow p -subgroup of A .

[Hint. Take $G = A_4$, $A = \{I, (123), (132)\}$, $H = \{I, (124), (142)\}$.

21. If H is a normal subgroup of a finite group G and P is a Sylow p -subgroup of H , show that $G = N_G(P)H$.

22. If a finite group G has an automorphism σ of prime order p ($p \mid o(G)$) then show that σ must fix at least one Sylow p -subgroup of G .

23. G is a finite group and P is a Sylow p -subgroup of G . H is a normal subgroup of G . If P is a normal subgroup of H then show that P is a normal subgroup of G .

24. Let M be the multiplicative group of all 3×3 non-singular matrices with entries in Z_p . Show that the set of all matrices of the form $\begin{pmatrix} 1 & a & b \\ 0 & 1 & c \\ 0 & 0 & 1 \end{pmatrix}$ where $a, b, c \in Z_p$ is a p -Sylow subgroup of M .

[Hint. Count the number of elements of M and number of elements of above type.]

3. Finite Abelian Groups

Let G be a finite Abelian group. As shown in Theorem 4.10, G is

expressible as a direct sum (product) of its Sylow subgroups. In particular if G is cyclic, we know that each subgroup of G is cyclic, hence G is expressible as direct sum (product) of Sylow subgroups each of which is cyclic. In this section we shall see that any finite Abelian group G is expressible as a direct sum (product) of a finite number of cyclic subgroups. For the sake of some convenience, we shall denote the binary composition multiplicatively.

Lemma 4.23. *Let a, b be two elements of finite order, of a group G such that $o(a)$ and $o(b)$ are co-prime and $ab=ba$. Then $o(ab)=o(a)o(b)$.*

Proof: Let $o(a)=m$ and $o(b)=n$. Let $H=\langle ab \rangle$ be the subgroup of G generated by ab . Since $(ab)^{mn}=a^{mn}b^{mn}=e$, we get $o(H)=o(ab) \mid mn$. Now $(ab)^m=a^mb^m=b^m \in H$, since $a^m=e$. As $(m, n)=1$, and $o(b)=n$, we have $o(b^m)=o(b)=n$. Now $b^m \in H \Rightarrow o(b^m) \mid o(H) \Rightarrow n \mid o(H)$. Similarly $m \mid o(H)$. Thus we have $mn \mid o(H)$. Hence $o(H)=mn=o(ab)$. This proves the Lemma. ■

Definition 4.24. (*Exponent of a group*). Let G be a finite group. If k is largest among the orders of all elements of G , then k is called the exponent of G , i.e. if there exists $a \in G$ such that $o(a)=k$ and no element of G has order exceeding k , then k is called the *exponent* of G .

Lemma 4.25. *The order of any element of a finite Abelian group G divides the exponent of G .*

Proof: Let k be the exponent of G , i.e. there exists an $a \in G$ such that $o(a)=k$ and for every $b \in G$, $o(b) \leq o(a)=k$. Suppose on the contrary there exists $b \in G$ such that $o(b) \nmid k$. This implies that there exists a prime number p such that for some positive integer t , $p^t \mid o(b)$ but $p^t \nmid o(a)$. This means that we can write $o(b)=p^\alpha q$, $o(a)=p^\beta s$ such that $(p, q)=1$, $(p, s)=1$, and $\beta < \alpha$. Let $b_o=b^q$, $a_o=a^{p^\beta}$. Then $o(b_o)=p^\alpha$, and $o(a_o)=s$. Lemma 4.23 gives $o(a_o b_o)=p^\alpha s > p^\beta s = o(a)$. This contradicts the assumption that $k=p^\beta s$ is the exponent of G . Hence the lemma follows. ■

EXAMPLE 10. S_3 is not commutative, the highest order of any element in S_3 is 3, i.e. the exponent of S_3 is 3. However (12) is an element of S_3 of order 2. Thus $o(12) \nmid 3$. This shows that the above lemma does not hold for non-Abelian groups.

Theorem 4.26. (*Fundamental Theorem of Finite Abelian Groups*). Every Abelian group G of order n , is an (internal) direct product

$$G=G_1 \times G_2 \times \dots \times G_t$$

where G_i are cyclic subgroups of order n_i such that $n_{i+1} \mid n_i$ and the integers n_i are uniquely determined. Further $n=n_1 n_2 \dots n_t$.

[Note. Whenever $o(G) > 1$, each G_i is taken with $o(G_i) > 1$.]

Proof: We prove the result by induction on $o(G)$. If $o(G)=1$,

the result holds trivially. Suppose $o(G)=n>1$ and that result holds for all Abelian groups of order $<o(G)$.

Let n_1 be the exponent of G and g_1 be an element of G of order n_1 .

Let $G_1=\langle g_1 \rangle$. If $G=G_1$, then G itself is cyclic. Let $G \neq G_1$. Then $1 < o(G/G_1) < n$. By the induction hypothesis

$$G/G_1 = H_2 \times H_3 \times \dots \times H_t$$

where each H_i is a cyclic subgroup of $\bar{G}=G/G_1$ of order $n_i > 1$ such that $n_{i+1} \mid n_i \forall i=2, 3, 4, \dots, t-1$, and $\frac{n}{n_1} = o(\bar{G}) = n_2 n_3 \dots n_t \dots (1)$

Now each $\bar{H}_i = H_i/G_1$ for some subgroup H_i of G containing G_1 . Choose $h_i \in H_i$ such that $h_i = h_i G_1$ is a generator of $\bar{H}_i$. Then $\bar{h}_i^{n_i} = \bar{e} = \bar{G}_1 \Rightarrow h_i^{n_i} \in G_1 = \langle g_1 \rangle \Rightarrow h_i^{n_i} = g_1^{m_i}$ for some m_i such that $1 \leq m_i \leq n_1$. Let $\alpha_i = (m_i, n_1)$. Then $m_i = \alpha_i \beta_i$ for some $\beta_i \geq 1$, $n_1 = \alpha_i \gamma_i^{m_i}$ for some $\gamma_i \geq 1$ and $(\beta_i, \gamma_i) = 1$.

Now $o(g_1) = n_1 = \alpha_i \gamma_i \Rightarrow o(g_1^{\alpha_i}) = \gamma_i$. As $(\gamma_i, \beta_i) = 1$, $o(g_1^{\alpha_i \beta_i}) = o(g_1^{\alpha_i})$. However $h_i^{n_i} = g_1^{\alpha_i \beta_i}$. So $o(h_i^{n_i}) = \gamma_i$. On the other hand $o(\bar{h}_i) \mid o(h_i) \Rightarrow n_i \mid o(h_i) \Rightarrow o(h_i^{n_i}) = \frac{o(h_i)}{n_i} \Rightarrow o(h_i) = o(h_i^{n_i}) n_i = \gamma_i n_i$

Since $o(h_i) \mid n_1$, the exponent of G and $n_1 = \alpha_i \gamma_i$, we have $\gamma_i n_i \mid \alpha_i \gamma_i$ or $n_i \mid \alpha_i \Rightarrow \alpha_i = n_i \delta_i$ for some $\delta_i \geq 1$. Then $h_i^{n_i} = g_1^{n_i \delta_i \beta_i}$.

Put $g_i = h_i (g_1)^{-\delta_i \beta_i} \forall i=2, 3, \dots, t-1$.

We see that $\bar{g}_i = g_i G_1 = h_i$, $g_i^{n_i} = h_i^{n_i} (g_1)^{-n_i \delta_i \beta_i} = e$.

This yields $o(g_i) = n_i$.

Define $H = \langle g_2 \rangle \times \langle g_3 \rangle \times \dots \times \langle g_t \rangle$. H is subgroup of G such that $o(H) \mid n_2 n_3 \dots n_t$.

Let $f: G \rightarrow G/G_1$ be the natural homomorphism.

Since $f(H) = \langle \bar{g}_2 \rangle \times \langle \bar{g}_3 \rangle \times \dots \times \langle \bar{g}_t \rangle$

$$= \bar{H}_2 \times \bar{H}_3 \times \bar{H}_4 \times \dots \times \bar{H}_t = G/G_1$$

and also $f(H) = (HG_1)/G_1$, we get $G = HG_1 = G_1 H$

$= \langle g_1 \rangle \times \langle g_2 \rangle \times \dots \times \langle g_t \rangle$. The fact that $o(G) = n = n_1 n_2 \dots n_t = o(g_1) o(g_2) \dots o(g_t)$ gives $G = \langle g_1 \rangle \times \langle g_2 \rangle \times \dots \times \langle g_t \rangle$.

Also $o(\langle g_i \rangle) = o(g_i) = n_i$ such that $n_{i+1} \mid n_i$.

Suppose $G = H_1 \times H_2 \times \dots \times H_t$ (3)

$$= K_1 \times K_2 \times \dots \times K_u$$
(4)

be two decompositions of G into internal direct product of cyclic subgroups, such that for $i=1, 2, \dots, t$, $j=1, 2, \dots, u$, $o(H_i) = n_i$, $o(K_j) = m_j$ and for $i \leq t-1$, $n_{i+1} \mid n_i$, for $j \leq u-1$, $m_{j+1} \mid m_j$. Consider $g \in G$, then (3) gives $g = h_1 h_2 \dots h_t$; $h_i \in H_i$.

Since $o(h_i) \mid o(H_i) = n_i$ and $n_i \mid n_1$, we get $h_{i n_1} = e$.

Consequently $g^{n_1} = h_1^{n_1} h_2^{n_1} \dots h_t^{n_1} = e$. Thus $o(g) \leq n_1 \forall g \in G$.

Further as H_1 is cyclic group of order n_1 , H_1 contains an element of order n_1 . Hence n_1 is the exponent of G . Similarly m_1 is the exponent of G . Thus $n_1 = m_1$. Suppose we have proved that $n_1 = m_1, n_2 = m_2, \dots, n_{i-1} = m_{i-1}$ for some i ; we shall prove that $n_i = m_i$. Suppose on the contrary $n_i \neq m_i$, and to be definite let $n_i > m_i$. Define $K = \{x^{m_i} \mid x \in G\}$. Since for any $x, y \in G$, $x^{m_i} y^{-m_i} = (xy^{-1})^{m_i} \in K$, we get that K is a subgroup of G . Suppose that for $k=1, 2, \dots, t$, $H_k = \langle a_k \rangle$ and for each $j=1, 2, \dots, u$, $K_j = \langle b_j \rangle$. Since for $j \geq i$, $o(K_j) = m_j \mid m_i$, $b_j^{m_i} = e$. Hence $K = \langle b_1^{m_i} \rangle \times \langle b_2^{m_i} \rangle \times \dots \times \langle b_{i-1}^{m_i} \rangle$.

Thus $o(K) = \frac{m_1}{m_i} \frac{m_2}{m_i} \dots \frac{m_{i-1}}{m_i}$, since $o(b_j) = m_j \nmid j$ (5)

Since also $G = \langle a_1 \rangle \times \dots \times \langle a_t \rangle$ [from (3)]
we have $K = \langle a_1^{m_i} \rangle \times \langle a_2^{m_i} \rangle \times \dots \times \langle a_t^{m_i} \rangle$.

Now $o(a_k^{m_i}) = \frac{n_k}{(m_i, n_k)} \nmid k=1, 2, \dots, t$.

Consequently $o(K) = \frac{n_1}{(m_i, n_1)} \frac{n_2}{(m_i, n_2)} \dots \frac{n_i}{(m_i, n_i)} \frac{n_{i+1}}{(m_i, n_{i+1})} \dots \frac{n_t}{(m_i, n_t)}$... (6)

Now $m_i \mid m_j$ and $m_j = n_j \nmid j < i$, by our hypothesis.

Thus $\frac{n_j}{(m_i, n_j)} = \frac{m_j}{(m_i, m_j)} = \frac{m_i}{m_i} \nmid j < i$.

Hence from (5) and (6) we obtain

$$1 = \frac{n_i}{(m_i, n_i)} \dots \frac{n_t}{(m_i, n_t)}. \quad \dots (7)$$

However $m_i < n_i$ gives $\frac{n_i}{(m_i, n_i)} > 1$. As a consequence, (7) cannot hold. This gives $m_i = n_i$. Hence by induction $m_i = n_i \nmid i$. However $n = n_1 n_2 \dots n_t = m_1 m_2 \dots m_u$. Thus we must also have $t = u$. This completes the theorem. ■

Corollary 4.27. Every finite Abelian p -group G is expressible as an internal direct product of cyclic groups

$$G = G_1 \times G_2 \times \dots \times G_r$$

such that $o(G) = p^{m_i} \nmid i=1, 2, \dots, r$ and $m_j \geq m_{j+1} \nmid j=1, 2, \dots, r-1$. Further p^{m_i} are uniquely determined.

Proof: By the above theorem, we can express

$$G = G_1 \times G_2 \times \dots \times G_r \quad \dots (1)$$

such that each G_i is a cyclic subgroup of order n_i and $n_{j+1} \mid n_j \nmid j \leq r-1$.

Since G is a p -group, each G_i is a p -group, we have $o(G) = p^m$ for some m , and $n_i = o(G_i) = p^{m_i} \nmid i$. Since $n_{j+1} \mid n_j$, we have $m_{j+1} \leq m_j$, $\nmid j \leq r-1$.

If $G = H_1 \times H_2 \times \dots \times H_s$

... (2)

is another decomposition of G into cyclic subgroups such that $o(H_i) = p^{u_i} \forall i = 1, 2, \dots, s$ and $u_i \geq u_{i+1} \forall j \leq s-1$, we see that $o(H_{j+1}) \mid o(H_j) \forall j \leq s-1$. Hence by the preceding theorem $r=s$ and $o(H_i) = o(G_i) \forall i = 1, 2, \dots, r$. This establishes the corollary. ■

REMARK 1. Let G be a finite Abelian p -group. As seen in the above corollary we can express $G = G_1 \times G_2 \times \dots \times G_r$ an internal direct product of cyclic p -subgroups G_i of order p^{n_i} . The numbers $p^{n_1}, p^{n_2}, \dots, p^{n_r}$ are uniquely determined by G , and are called the *invariants* of G . Let G' be another finite Abelian group. Clearly if $G \cong G'$ then G and G' must have same set of invariants. Conversely if G' has the same set of invariants as G , then by definition of invariants, we get $G' = G'_1 \times G'_2 \times \dots \times G'_r$, such that G'_i are cyclic groups of order p^{n_i} . Since any two cyclic groups of same order are isomorphic, we get $G_i \cong G'_i$. This, then gives that $G \cong G'$. Hence two finite Abelian p -groups are isomorphic if and only if they have the same set of invariants. Thus for example if G is of order p^3 , the possible sets of invariants of G are $\{p, p, p\}$, $\{p^2, p\}$ and $\{p^3\}$. Consequently there exist (within isomorphism) only three Abelian groups of order p^3 . Again if order of G is p^2 , then the possible sets of invariants of G are $\{p^2\}$, $\{p, p\}$, so that number of non isomorphic Abelian groups of order p^2 is two.

REMARK 2. Consider any finite Abelian group. We know that $G = S(p_1) \times S(p_2) \times \dots \times S(p_t)$ [Theorem 4.10]. Since each of $S(p_i)$ is a p_i -group, $S(p_i)$ can be expressed as finite direct product of cyclic p_i -groups. Consequently G itself can be expressed as a finite direct product of cyclic groups of prime power orders. As in the above corollary, one can see that the orders of these cyclic groups are uniquely determined by G itself. The set of these orders can be again termed as the set of invariants of G . Again two finite Abelian groups are isomorphic if and only if they have same set of invariants.

WORKED-OUT EXERCISES

Exercise 1. Let G be an abelian group and $a, b \in G$ be of order m and n respectively. Show that there exists an element c of G such that $o(c) = k$ where k is LCM of m and n .

Solution. Firstly suppose that $(m, n) = 1$ then $k = mn$.

Put $c = ab$. Clearly $c^{mn} = (a^m)^n (b^n)^m = e \Rightarrow o(c)$ is finite and $\leq mn$. Let $o(c) = t$. Then $e = c^t = a^t b^t \Rightarrow a^t = b^{-t} \Rightarrow e = a^{mt} = b^{-mt} \Rightarrow b^{mt} = e \Rightarrow n \mid mt \Rightarrow n \mid t$ as $(m, n) = 1$.

Similarly it can be proved that $m \mid t$. Again as $(m, n) = 1$, we get that $mn \mid t$ i.e. $mn \leq t$. Hence $o(c) = mn$.

Now suppose that $(m, n) > 1$. We can put $m = p_1^{\alpha_1} p_2^{\alpha_2} \dots p_u^{\alpha_u}$ and $n = p_1^{\beta_1} p_2^{\beta_2} \dots p_u^{\beta_u}$ where p_i 's are distinct prime numbers and α_i, β_i are non-negative integers such that $\alpha_i \geq \beta_i \forall 1 \leq i \leq t$ and $\alpha_j < \beta_j \forall t+1 \leq j \leq u$. Then $k = p_1^{\alpha_1} p_2^{\alpha_2} \dots p_t^{\alpha_t} p_{t+1}^{\beta_{t+1}} \dots p_u^{\beta_u}$.

Put $g = a p_1^{\beta_{t+1}} \dots p_k^{\alpha_k}$ and $h = b p_1^{\beta_1} p_2^{\beta_2} \dots p_t^{\beta_t}$.

Therefore, $c(g) = p_1^{\alpha_1} p_2^{\alpha_2} \dots p_t^{\alpha_t}$ and $o(h) = p_{t+1}^{\beta_{t+1}} \dots p_u^{\beta_u}$. Clearly $o(g)$ and $o(h)$ are co-prime, hence by first case $c = gh$ has order equal to $p_1^{\alpha_1} p_2^{\alpha_2} \dots p_t^{\alpha_t} p_{t+1}^{\beta_{t+1}} \dots p_u^{\beta_u}$ i.e. $o(c) = k$.

Exercise 2. If n is the exponent of an abelian group G then for every $a \in G$, $c(a)$ divides n .

Solution. Let $o(a) = m$, if m does not divide n then $[m, n] = \text{LCM of } m \text{ and } n$, is strictly greater than n . By Exercise 1 above, there exists an element $c \in G$ such that $o(c) = [m, n] > n$. This is a contradiction to the definition of exponent.

[Note. This provides another proof of Lemma 4.25.]

Exercise 3. Let G be a finite abelian group in which the number of solutions in G of the equation $x^n = e$ is at most n for every positive integer n . Prove G must be a cyclic group.

Solution. Let the order of the given group G be m . Let n_1 be the exponent of G . Since for any $a \in G$, $o(a) \mid n_1$, $a^{n_1} = e$ for all $a \in G$. Hence $x^{n_1} = e$ has m solutions. Thus $m \leq n_1$ as $x^{n_1} = e$ has at most n_1 solutions. As exponent of G is n_1 , there exists an element $b \in G$ such that $o(b) = n_1$. But $o(b) \mid o(G) \Rightarrow n_1 \mid m \Rightarrow n_1 \leq m$. Hence $m = n_1$ i.e. $o(b) = m = o(G)$.

As a consequence G is a cyclic group generated by b .

PROBLEMS

1. Find the invariants of a cyclic group of order n .

[Hint. If $n = p_1^{\alpha_1} p_2^{\alpha_2} \dots p_t^{\alpha_t}$ and G is a cyclic group of order n , then $G = G_1 \times G_2 \times \dots \times G_t$ where G_i is the cyclic subgroup of G of order $p_i^{\alpha_i}$.]

2. Using the Fundamental Theorem of Finite Abelian Groups, prove the following: If G is a finite Abelian group of exponent n , then no proper subgroup of G contains all the elements of order n .

3. Give a detailed proof of the fact that two finite Abelian groups are isomorphic if and only if they have same sets of invariants.

4. An Abelian group has invariants p, p , where p is a prime. How many subgroups of order p it has? [Ans. $p+1$]
5. For $n=20$, how many non-isomorphic Abelian groups of order n exist? [Ans. Two]
6. An Abelian group G of order p^3 , p a prime, has invariants p^2, p . Determine all its subgroups of order p .
7. For any positive integer n , and for any prime p , prove that the number of non-isomorphic Abelian p -groups of order p^n is equal to the number of partitions of n .

SOLVABLE GROUPS AND JORDAN-HOLDER THEOREM

1. Generators of a Subgroup and Derived Subgroups

We have already discussed earlier, the concept of a generator of a cyclic group, here we attempt to generalize the same concept to arbitrary groups.

Definition 5.1. Let S be a subset of a group G . A subgroup H of G is said to be *generated* by S if it satisfies the following conditions:

(1) $S \subseteq H$.

(2) If K is any subgroup of G such that $S \subseteq K$, then $H \subseteq K$.

We denote the subgroup generated by S by $\langle S \rangle$.

REMARK. Suppose F is the family of all those subgroups of G which contain S . This family is not void, since $S \subseteq G$ and so $G \in F$. Let H be the intersection of all members of F . Clearly if K is any subgroup of G containing S , then $H \subseteq K$; further $S \subseteq H$, since $S \subseteq K \forall K \in F$. Hence H is the subgroup of G generated by S . So, alternatively, we can say that a subgroup H of G is said to be generated by S if H is the intersection of all those subgroups of G which contain S .

Now for a subset S of a group G it is possible that G itself is the subgroup generated by S . Then S is called a *set of generators* of G . For example if $G = \langle a \rangle$ is a cyclic group, then $S = \{a\}$ is a set of generators of G . However, here S is a singleton.

Theorem 5.2. If S is any non-void subset of a group G , then the subgroup $\langle S \rangle$ of G generated by S is the set of all finite products of the form $a_1 a_2 \dots a_n$ where for each i , either $a_i \in S$ or $a_i^{-1} \in S$.

Proof: Let H be the set of all finite products of the form $a_1 a_2 \dots a_n$; a_i or $a_i^{-1} \in S$, n any positive integer. Consider $x = a_1 a_2 \dots a_n$, $y = b_1 b_2 \dots b_m$ in H . Then $xy = a_1 a_2 \dots a_n b_1 b_2 \dots b_m$ is a product of finite number of elements a_i, b_j such that either the factor or its inverse is in S , consequently $xy \in H$. Further $x^{-1} = a_n^{-1} a_{n-1}^{-1} \dots a_1^{-1}$. Consider any a_i^{-1} . Since a_i or a_i^{-1} is in S , and $a_i = (a_i^{-1})^{-1}$, we see

that either a_i^{-1} is $(a_i^{-1})^{-1}$ is in S , hence $x^{-1} \in H$. This proves that H is a subgroup of G . Clearly $S \subseteq H$. Consider any subgroup K of G containing S . Then for each $a \in S$, $a \in K$ and hence $a^{-1} \in K$. Thus if $x = a_1 a_2 \dots a_n : a_i \in S$ or $a_i^{-1} \in S$, is any element of H , then $x \in K$, since $a_i \in K \forall i$. Hence $H \subseteq K$. This proves that H is the subgroup of G generated by S . ■

EXAMPLE 1. Let A be the set of all transpositions in the symmetric group S_n . Since every $f \in S_n$ is a product of transpositions, and so, of members of A , we get that A generates S_n .

EXAMPLE 2. $S = \{1\}$ is a generator of the additive group $\mathbb{Z}$ of integers.

EXAMPLE 3. Consider $S_3 = \{I, (12), (13), (23), (123), (132)\}$. Let $a = (12)$, $b = (123)$. Then $b^2 = (132)$, $b^3 = I$, $ba = (13)$, $ab = (23)$. Hence, we see that every member of S_3 is expressible as a product of members of $A = \{a, b\}$. Consequently A generates S_3 .

EXAMPLE 4. Let $G = \{\pm 1, \pm i, \pm j, \pm k\}$ be the group of quaternions and $A = \{i, j\}$. Let H be the subgroup of G generated by A . Then $i \in H, j \in H, ij = k \in H$. Also $-i = i^3 \in H$, $-j = j^3 \in H$, $-k = k^3 \in H$, $-1 = i^2 \in H$, $1 = i^4 \in H$. This proves that H contains all the members of G . Consequently $H = G$ and A generates G .

Definition 5.3. Let a, b be any two elements of a group G . The commutator of a and b is the element $a^{-1}b^{-1}ab$. The commutator subgroup of G denoted by G' , is the subgroup of G generated by the set of all commutators of G .

The commutator subgroup is also called a *first derived subgroup* or simply a *derived subgroup*.

Theorem 5.4. Let G be a group and G' be its commutator subgroup then

(1) G' is a normal subgroup of G .

(2) For any normal subgroup H of G , G/H is an Abelian group if and only if H contains G' .

Proof: (1) Let $a, b \in G$, since $(a^{-1}b^{-1}ab)^{-1} = b^{-1}a^{-1}ba$ is again a commutator, we see that each element of G' is a product of finite number of commutators (Theorem 5.2). Consider $x \in G'$, then $x = g_1 g_2 \dots g_t$ where for each $i = 1, 2, \dots, t$, g_i is a commutator, so that $g_i = a_i^{-1} b_i^{-1} a_i b_i$ for some $a_i, b_i \in G$. Now for any $a \in G$, $a^{-1} x a = (a^{-1} g_1 a) (a^{-1} g_2 a) \dots (a^{-1} g_t a)$

Further $a^{-1} g_i a = a^{-1} a_i^{-1} b_i^{-1} a_i b_i a$

$$= (a^{-1} a_i a)^{-1} (a^{-1} b_i a)^{-1} (a^{-1} a_i a) (a^{-1} b_i a)$$

$$= c^{-1} d^{-1} c d; \text{ where } c = a^{-1} a_i a, d = a^{-1} b_i a$$

gives $a^{-1} g_i a$ is again a commutator. Hence $a^{-1} x a$ is a product of commutators; by definition $a^{-1} x a \in G'$. This proves the result.

(2) Consider $a, b \in G$. Let G/H be Abelian.

This $\Rightarrow (aH)(bH) = (bH)(aH)$

$$\Rightarrow abH = baH \Rightarrow a^{-1}b^{-1}ab = (ba)^{-1}(ab) \in H.$$

Thus H contains every commutator $a^{-1}b^{-1}ab$. Consequently as G' is generated by all the commutators, $G' \subseteq H$.

Conversely let $G' \subseteq H$. Then $a^{-1}b^{-1}ab \in G'$ gives $a^{-1}b^{-1}ab \in H \Rightarrow abH = baH \Rightarrow (aH)(bH) = (bH)(aH)$. This establishes that G/H is Abelian. ■

Taking $H = (e)$, we get.

Corollary 5.5. A group G is commutative if and only if $G' = (e)$. ■

PROBLEMS

1. Prove that in $S_n (n \geq 3)$, the set S of all 3-cycles generate A_n .

[Hint. Every member of A_n is a product of even number of transpositions. For any two transpositions (ij) and (kl) in S_n we have $(ij)(kl) = (ijl)$ if $k=j$, and when $k \neq i, j, l \neq i, j$, then $(ij)(kl) = (ij)(jk)(kl) = (ijk)(klj)$.]

2. From the above problem conclude that the set of all 3-cycles of the form $(1ij)$ generate A_n .

[Hint. Show that any 3-cycle is a product of 3-cycles of the form $(1ij)$.]

3. For any endomorphism σ of a group G , prove that the image of a commutator $a^{-1}b^{-1}ab$ under σ , is again a commutator. Deduce that G' is a characteristic subgroup of G .

4. For any group G , prove that

$$G' = \{a_1 a_2 \dots a_n a_1^{-1} a_2^{-1} a_3^{-1} \dots a_n^{-1} ; a_i \in G, n \geq 2\}.$$

5. For $G = S_3$, prove that $G' = A_3$.

6. For any simple non-commutative group G , prove that $G' = G$. Deduce that for any $n \geq 5$, $S'_n = A_n$.

[Hint. A_n is simple, non-commutative and $S'_n \subseteq A_n$.]

7. If H and K are subgroups of a group G , then show $H \subseteq K \Rightarrow H' \subseteq K'$.

8. If the commutator subgroup of a group G is of order m then show that each element of G has at most m conjugates.

9. Let G be a group whose centre Z has index n in G , show that G has at most n^2 different commutators.

10. N is a normal subgroup of a group G and $N \cap G' = \{e\}$. Show that $N \subseteq Z(G)$.

2. Normal Series, Solvable Groups and Jordan-Holder Theorem

Definition 5.6. (Subnormal Series): A sequence of subgroups

$$G = G_0 \supseteq G_1 \supseteq G_2 \supseteq \dots \supseteq G_\lambda = (e) \quad \dots (1)$$

of a group G is called a *subnormal series* of G , if G_{i+1} is a normal subgroup of $G_i \neq i=0, 1, \dots, \lambda-1$. The factor groups G_i/G_{i+1} are called the *factor groups* of the subnormal series. Further if each G_i is a normal subgroup of G itself, then the series is said to be a *normal series* of G .

Note. By the statement " $N=\{G_0, G_1, \dots, G_\lambda\}$ is a subnormal series of a group G " we mean the sequence (1) of subgroups G_i such that G_{i+1} is a normal subgroup of G_i for all $i=0, 1, \dots, \lambda-1$.

Definition 5.6 (a). (Refinement).

Let $N=\{G_0, G_1, \dots, G_\lambda\}$ be a subnormal series of a group G then a subnormal series $M=\{H_0, H_1, \dots, H_\mu\}$ is said to be a *refinement* of N if $N \subseteq M$.

Definition 5.6 (b). (Proper Refinement)

A refinement M of N is said to be a *proper refinement*, if $N \subset M$ but $N \neq M$.

REMARK. Most of the authors use the term normal series, in stead of "subnormal series".

It is possible that for some i , $G_i = G_{i+1}$ in (1). The number of distinct members of (1) different from G is called the *length* of the subnormal series. The subnormal series is said to be *redundant* if for some $i=0, 1, 2, \dots, \lambda-1$, $G_i = G_{i+1}$, otherwise it is said to be *irredundant*. One can always construct an irredundant subnormal series from a redundant one by deleting G_{i+1} whenever for some i , $G_{i+1} = G_i$. Thus in case the subnormal series (1) is irredundant, all G_i 's are distinct; consequently its length is λ itself.

Definition 5.7. A group G is said to be *solvable* if it has a subnormal series

$$G = G_0 \supseteq G_1 \supseteq G_2 \supseteq \dots \supseteq G_s = (e)$$

such that each of its factor group G_i/G_{i+1} is an Abelian group. The above series, then is referred to as a *solvable series* of G .

EXAMPLE 5. Any Abelian group G is solvable. Now $G > (e)$ is a normal series of G and its only factor group is $G/(e)$, which being isomorphic to G , is commutative.

EXAMPLE 6. Consider the series $S_3 > A_3 > (I)$. Its factor groups are S_3/A_3 and A_3 which are of orders 2 and 3 respectively. Since any group of prime order is commutative, it follows that both the factor groups are commutative. Hence S_3 is solvable.

EXAMPLE 7. Consider $S_4 > A_4 > V_4 > \{I\}$, a subnormal series of S_4 , where $V_4 = \{I, (12)(34), (13)(24), (14)(23)\}$. Its factor groups are S_4/A_4 , A_4/V_4 , and V_4 which are of orders 2, 3 and 4 respectively. Since every group of order 4 is Abelian, it follows that all the factor

groups are Abelian, hence S_4 is solvable.

Theorem 5.8. Any subgroup H of a solvable group G is solvable.

Proof: Let $G = G_0 \supseteq G_1 \supseteq G_2 \supseteq \dots \supseteq G_n = (e)$

be a solvable series for G . We claim that

$$H = H_0 \supseteq (H \cap G_1) \supseteq (H \cap G_2) \supseteq \dots \supseteq H \cap G_n = (e) \quad \dots (1)$$

is a solvable series for H . Since for $i=0, 1, \dots, n-1$, G_{i+1} is normal in G_i , we get $H_{i+1} = H \cap G_{i+1}$ is normal in $H_i = H \cap G_i$. Define a mapping $f: H_i \rightarrow G_i/G_{i+1}$ such that $f(x) = xG_{i+1} \forall x \in H_i$. f is a homomorphism. Further

$$x \in \text{Ker } f \Leftrightarrow xG_{i+1} = G_{i+1}$$

$$\Leftrightarrow x \in G_{i+1}$$

$$\Leftrightarrow x \in H \cap G_{i+1} \text{ since } x \in H_i \subseteq H.$$

This yields that $\text{Ker } f = H \cap G_{i+1} = H_{i+1}$ hence by the Fundamental Theorem of Homomorphism $H_i/H_{i+1} \cong f(H_i)$. As $f(H_i)$ is a subgroup of G_i/G_{i+1} and G_i/G_{i+1} is Abelian, $f(H_i)$ is also Abelian, Consequently H_i/H_{i+1} is Abelian. This proves that (1) is a solvable series for H . Hence H is a solvable group. ■

Theorem 5.9. If H is a normal subgroup of a solvable group G , then G/H is also solvable.

Proof: Let $G = G_0 \supseteq G_1 \supseteq G_2 \supseteq \dots \supseteq G_n = (e)$

be a solvable series for G . Consider the series

$$\frac{G}{H} = G_0/H \supseteq G_1H/H \supseteq G_2H/H \supseteq \dots \supseteq G_nH/H = (\bar{e}). \quad \dots (1)$$

For each $i=0, 1, 2, \dots, n-1$, G_{i+1} is normal in G_i . Consider any $x \in G_iH$ then $x = gh$, for some $g \in G_i$, $h \in H$. Thus $xG_{i+1}H = g h G_{i+1}H$
 $= gh H G_{i+1}$ (since for any subgroup K of G , $HK = KH$)
 $= g G_{i+1} H$, since $h H = H$.
 $= G_{i+1} g H = G_{i+1} gh H$, since G_{i+1} is normal in G_i and $g \in G_i$

$$= G_{i+1} Hgh = G_{i+1} Hx.$$

This proves that $G_{i+1}H$ is a normal subgroup G_iH .

$$\text{Hence } G_iH/G_{i+1}H_{i+1} \cong \frac{G_iH/H}{G_{i+1}H/H} \quad \dots (2)$$

Now define $f: G_i \rightarrow G_iH/G_{i+1}H$

such that $f(x) = G_{i+1}Hx \forall x \in G_i$, then f is a homomorphism. As $G_iH = HG_i$, given $y \in G_iH$, we can write

$$y = hg \text{ for some } h \in H, g \in G_i; \text{ then}$$

$$G_{i+1}H y = G_{i+1}H hg = G_{i+1}H g = f(g)$$

This shows that f is also onto, clearly $G_{i+1} \subseteq \text{Ker } f$, and hence induces a homomorphism

$$\bar{f}: G_i/G_{i+1} \rightarrow G_iH/G_{i+1}H$$

such that $\bar{f}(G_{i+1}x) = G_{i+1}Hx \forall x \in G_i$.

This σ is also onto. Thus $G_i H / G_{i+1} H$ is a homomorphic image of the Abelian group G_i / G_{i+1} . So that it must be itself Abelian. Consequently each factor group of the subnormal series (1) is Abelian. This proves that G/H is solvable. ■

Theorem 5.10. Let H be a normal subgroup of a group G . If both H and G/H are solvable then G is solvable.

Proof: Let $G/H = G_0/H \supseteq G_1/H \supseteq \dots \supseteq G_{t-1}/H \supseteq G_t/H = (e) \dots (1)$ be a solvable series for G/H . Here each G_i is a subgroup of G containing H . Since G_{i+1}/H is normal in G_i/H , each G_{i+1} is normal in G_i . Further $G_t/H = e$ implies $G_t = H$. Now let

$H = H_0 \supseteq H_1 \supseteq H_2 \supseteq \dots \supseteq H_u = (e)$ be a solvable series of H . $\dots (2)$ Then $G = G_0 \supseteq G_1 \supseteq G_2 \supseteq \dots \supseteq G_{t-1} \supseteq H \supseteq H_1 \supseteq H_2 \supseteq \dots \supseteq H_u = (e)$ is a solvable series for G . This proves that G is a solvable group. ■

Corollary 5.11. If two groups H and K are solvable, then $H \times K$ is solvable.

Proof: Let $G = H \times K$. Then $G/H \cong K$. Since K is solvable, we have G/H is solvable. Also we are given that H is solvable. Hence G is solvable by the above theorem. ■

Corollary 5.12. Any finite p -group is solvable.

Proof: Let G be a finite p -group (p is a prime number) then $o(G) = p^n$, for some $n \geq 0$. If $n=1$, G is Abelian and so it is solvable. So let $n > 1$, we apply induction on n . Suppose the result holds for groups of order p^m with $m < n$. Now $Z(G)$, the centre of G is not equal to (e) . Consequently $o[Z(G)] = p^k$ for some $k \geq 1$. This yields that $G/Z(G)$ is a group of order p^{n-k} with $n-k < n$. Hence by the induction hypothesis $G/Z(G)$ is solvable. Also $Z(G)$ being Abelian, is solvable. Hence, taking $H = Z(G)$, we see from the above theorem that G is solvable. ■

In the last section we introduced the concept of a commutator subgroup. We now establish a criterion for the solvability of a group by means of commutator subgroups.

Definition 5.13. Let G be a group, for every non-negative integer n , define $G^{(n)}$ inductively as follows.

$$G^{(0)} = G$$

$G^{(n+1)} = (G^{(n)})'$, the commutator subgroup of $G^{(n)}$. $G^{(n)}$ is called n th commutator subgroup or n th derived subgroup of G .

Now we establish the following theorem.

Theorem 5.14. A group G is solvable if and only if $G^{(n)} = (e)$ for some non-negative integer n .

Proof: Let G be solvable and let

$$G = G_0 \supseteq G_1 \supseteq G_2 \supseteq \dots \supseteq G_n = (e)$$

be a solvable series. We prove inductively that $G^{(k)} \subseteq G_k \forall k$.

Clearly $G^{(0)} = G = G_0$.

Suppose for some k , $G^{(k)} \subseteq G_k$, this gives

$$G^{(k+1)} = [G^{(k)}]' \subseteq G_k'.$$

As G_k/G_{k+1} is Abelian, we have $G_k' \subseteq G_{k+1}$. (Theorem 5.4.)

Consequently $G^{(k+1)} \subseteq G_{k+1}$. Hence by induction

$G^{(k)} \subseteq G_k \forall k=0, 1, 2, \dots, n$. In particular $G^{(n)} \subseteq G_n$. Hence $G^{(n)} = (e)$ since $G_n = (e)$.

Conversely let $G^{(n)} = (e)$ for some n ,

then $G = G^0 \supseteq G^{(1)} \supseteq G^{(2)} \supseteq \dots \supseteq G^{(n)} = (e)$

is a subnormal series for G , such that $G^{(i)}/G^{(i+1)} = G^{(i)}/[G^{(i)}]'$ is Abelian since for any group H , H/H' is always Abelian, hence G is solvable. This establishes the theorem. ■

Corollary 5.15. S_n is not solvable for $n \geq 5$.

Proof: Now A_n is a simple non-commutative group, so $A_n' \neq (I)$. However A_n is simple, so its only normal subgroups are A_n and (I) . Consequently $A_n' = A_n \Rightarrow A_n^{(2)} = (A_n')' = A_n' = A_n$. In general $A_n^{(k)} = A_n \forall$ positive integers k , thus $A_n^{(k)} \neq (e) \forall k$. Hence A_n is not solvable.

As now A_n is a subgroup of S_n , S_n is also not solvable for $n \geq 5$, since subgroup of a solvable group is solvable. ■

Definition 5.16. (Maximal normal subgroup). A normal subgroup H of a group G is said to be a maximal normal subgroup if $H \neq G$ and there does not exist any normal subgroup K of G such that $H < K < G$.

For a normal subgroup H of G , we know that any subgroup of G/H is of the form K/H , where K is a subgroup of G containing H . Further K/H is normal in G/H if and only if K is normal in G . K/H is a proper subgroup of G/H if and only if $H < K < G$. Thus H is a maximal normal subgroup of G if and only if $H \neq G$, and G/H has no proper normal subgroups i.e. G/H is simple.

EXAMPLE 8. Let $G = \langle a \rangle$ be a cyclic group of order 10. Then $H = \langle a^2 \rangle$ is subgroup of G such that G/H is of order 2. Thus G/H is simple and H is maximal normal subgroup of G .

EXAMPLE 9. If G is a simple group then (e) is a maximal normal subgroup of G since G contains no proper normal subgroup.

EXAMPLE 10. In S_n ($n > 2$), A_n is a maximal normal subgroup.

We recall the definition of refinement [Definition 5.6 (a)] and restate it as under:

Definition 5.17. Let $G = G_0 \supseteq G_1 \supseteq \dots \supseteq G_\lambda = (e)$ be a subnormal series of G . A subnormal series $G = H_0 \supseteq H_1 \supseteq H_2 \supseteq \dots \supseteq H_\mu = (e)$ is called a refinement of the former series if every G_i is one of the H_j 's.

Definition 5.18. An irredundant subnormal series

$$G = G_0 > G_1 > G_2 > \dots > G_t = (e)$$

of a group G is said to be a *composition series* of G , if each of its factor groups G_i/G_{i+1} is a simple group, i.e. G_{i+1} is a maximal normal subgroup of G_i .

EXAMPLE 11. Let $G = \langle a \rangle$ be a cyclic group of order 24, then $o(a) = 24$. Consider two series

$$G = \langle a \rangle > \langle a^2 \rangle > \langle a^6 \rangle > \langle a^{12} \rangle > (e) \dots (1)$$

$$G = \langle a \rangle > \langle a^3 \rangle > \langle a^8 \rangle > \langle a^{12} \rangle > (e) \dots (2)$$

These are two normal series of G . The factor groups of (1) are $\langle a \rangle / \langle a^2 \rangle$, $\langle a^2 \rangle / \langle a^6 \rangle$, $\langle a^6 \rangle / \langle a^{12} \rangle$ and $\langle a^{12} \rangle / (e)$ which are of orders 2, 3, 2, 2 respectively. Since these are of prime

orders, they are all simple. Similarly the factor groups of (2) are

$\langle a \rangle / \langle a^3 \rangle$, $\langle a^3 \rangle / \langle a^8 \rangle$, $\langle a^8 \rangle / \langle a^{12} \rangle$, $\langle a^{12} \rangle / (e)$, which are of orders 3, 2, 2, 2 respectively.

These are again simple. So (1) and (2) both are composition series of G . One can see that both of these series are of same length viz. 4. Further as two cyclic groups of same order are isomorphic, we have

$$\langle a \rangle / \langle a^2 \rangle \cong \langle a^3 \rangle / \langle a^6 \rangle, \langle a^2 \rangle / \langle a^6 \rangle \cong \langle a \rangle / \langle a^3 \rangle, \\ \langle a^6 \rangle / \langle a^{12} \rangle \cong \langle a^8 \rangle / \langle a^{12} \rangle, \langle a^{12} \rangle / (e) \cong (e) / (e).$$

Thus there is a 1-1 correspondence between the factor groups of (1) and those of (2) such that the corresponding factor groups are isomorphic.

EXAMPLE 12. Let $G = S_4$. Consider two subnormal series

$$S_4 > A_4 > V_4 > (I) \text{ (see Example 9)} \dots (1)$$

$$\text{and } S_4 > A_4 > V_4 > A > (I) \dots (2)$$

$$\text{where } A = \{I, (12), (34)\}.$$

$V_4 \cong V_4 / (I)$ is a factor group of (1). As V_4 is a group of order 4, it is not simple (since any simple commutative group is of prime order). The factor groups of (2) viz. S_4/A_4 , A_4/V_4 , V_4/A , $A/(I)$ are of orders 2, 3, 2 and 2 respectively. Hence all these are simple and (2) is a composition series of S_4 .

Henceforth the main purpose is to prove the Jordan-Hölder theorem. To achieve this objective, firstly we introduce a few notions and prove some results.

Definition 5.19. Two subnormal series

$$G = G_0 \supseteq G_1 \supseteq G_2 \supseteq \dots \supseteq G_s = (e) \dots (1)$$

$$\text{and } G = G_0' \supseteq G_1' \supseteq G_2' \supseteq \dots \supseteq G_t' = (e) \dots (2)$$

of a group G are said to be *equivalent* or *isomorphic* if there exists a 1-1 correspondence between factors of (1) and those of (2) such that the corresponding factor groups are isomorphic.

Since the number of factor groups of (1) and (2) are s and t , if the

series are equivalent, then by definition they must have same number of factor groups. Consequently $s=t$. Thus in Example 11 both the series are equivalent or isomorphic.

Following lemma is analogous to Second Theorem of Isomorphism.

Lemma 5.20. Let H and K be two subgroups of G such that $kH = Hk$ for every $k \in K$. Then HK is a subgroup of G , H is a normal subgroup of HK , $H \cap K$ is a normal subgroup of K and

$$\frac{HK}{H} \cong \frac{K}{H \cap K}.$$

Proof: Since $kH = Hk \forall k \in K$ we have $HK = KH$. Consequently HK is a subgroup of G (Theorem 2.39). Consider $x \in HK$, $x = hk = kh_1$ for some $h, h_1 \in H, k \in K$. Then, $Hx = Hhk = Hk = kH = k(h_1H) = kh_1H = xH$ i.e. $Hx = xH \forall x \in HK$. This yields that H is a normal subgroup of HK . It can be easily checked that $H \cap K$ is a normal subgroup of K . By the Second Theorem of Isomorphism $HK/H \cong K/H \cap K$.

This establishes the result. ■

Theorem 5.21. (Zassenhaus) Let B and C be any two subgroups of a group G , B_0 and C_0 be normal subgroups of B and C respectively, then

$$\frac{B_0(B \cap C)}{B_0(B \cap C_0)} \cong \frac{C_0(C \cap B)}{C_0(C \cap B_0)}.$$

REMARK. Note the symmetry, indeed the RHS of the isomorphism is obtained from LHS by interchanging B and C . This theorem is also known as *Butterfly theorem*.

Proof: Let $K = B \cap C$, $H = B_0(B \cap C_0)$. Since B_0 is a normal subgroup of B , $bB_0 = B_0b \forall b \in B$, in particular as $K \subseteq B$ we have $kB_0 = B_0k \forall k \in K$. Also C_0 is normal in C yields that $B \cap C_0$ is normal in $B \cap C$. Consequently $k(B \cap C_0) = (B \cap C_0)k \forall k \in K$. Hence

$$\begin{aligned} Hk &= B_0(B \cap C_0)k = B_0[k(B \cap C_0)] \\ &= (B_0k)(B \cap C_0) = (kB_0)(B \cap C_0) \\ &= k[B_0(B \cap C_0)] = kH \forall k \in K. \end{aligned}$$

$$\text{Therefore } \frac{HK}{H} \cong \frac{K}{H \cap K}. \quad \dots (1)$$

Now $HK = B_0(B \cap C_0)(B \cap C) = B_0(B \cap C)$, since $B \cap C_0 \subseteq B \cap C$, $(B \cap C_0)(B \cap C) = B \cap C$. (Indeed for any group G , $gG = G \forall g \in G$).

Further $y \in H \cap K \Rightarrow y \in H, y \in K \Rightarrow y = b_0b, y = d$ for some $b_0 \in B_0, b \in B \cap C_0, d \in B \cap C$.

$\Rightarrow b_0b = d \Rightarrow b_0 = db^{-1} \in B_0 \cap C = C \cap B_0$ as $d \in C$ and $b^{-1} \in B \cap C_0 \subseteq C_0 \subseteq C$.

Thus $y = b_0b \in (C \cap B_0)(B \cap C_0)$.

This yields that $H \cap K \subseteq (C \cap B_0)(B \cap C_0)$. Since $C \cap B_0 \subseteq B \cap C$, $B \cap C_0 \subseteq B \cap C$, $(C \cap B_0)(B \cap C_0) \subseteq B \cap C = K$.

Also $C \cap B_0 \subseteq B_0$ gives $(C \cap B_0)(B \cap C_0) \subseteq B_0(B \cap C_0) = H$. Thus we get

$(C \cap B_0)(B \cap C_0) \subseteq H \cap K$ Consequently $H \cap K = (C \cap B_0)(B \cap C_0)$.

Then putting $HK = B_0(B \cap C)$, $H \cap K = (C \cap B_0)(B \cap C_0)$ in (1), we get that

$$\frac{B_0(B \cap C)}{B_0(B \cap C_0)} \cong \frac{B \cap C}{(C \cap B_0)(B \cap C_0)}. \quad \dots (2)$$

Interchanging the roles of B and C we get

$$\frac{C_0(C \cap B)}{C_0(C \cap B_0)} \cong \frac{(C \cap B)}{(B \cap C_0)(C \cap B_0)}. \quad \dots (3)$$

By using the fact that $B \cap C_0$ is a normal subgroup of $B \cap C$ and $C \cap B_0 \subseteq B \cap C$, we get

$$(B \cap C_0)(C \cap B_0) = (C \cap B_0)(B \cap C_0).$$

$$\text{Hence } \frac{B \cap C}{(C \cap B_0)(B \cap C_0)} = \frac{C \cap B}{(B \cap C_0)(C \cap B_0)}.$$

Then from (2) and (3), it follows that

$$\frac{B_0(B \cap C)}{B_0(B \cap C_0)} \cong \frac{C_0(C \cap B)}{C_0(C \cap B_0)}.$$

This proves the theorem. ■

Theorem 5.22. (Schreier's refinement theorem). Any two subnormal series of a group have equivalent refinements.

Proof: Consider the subnormal series

$$G = G_0 \supseteq G_1 \supseteq G_2 \supseteq \dots \supseteq G_s = (e) \quad \dots (1)$$

$$G = H_0 \supseteq H_1 \supseteq H_2 \supseteq \dots \supseteq H_t = (e) \quad \dots (2)$$

of a group G . Since for any i, j, k and l , G_{i+1} is normal subgroup of G_i and H_{k+1} is a normal subgroup of H_k , we get

$$G_{i,j} = G_{i+1}(G_i \cap H_j), \quad (i=0, 1, 2, \dots, s-1; j=0, 1, 2, \dots, t) \quad \dots (3)$$

$$H_{k,l} = H_{k+1}(H_k \cap G_l), \quad (k=0, 1, 2, \dots, t-1; l=0, 1, 2, \dots, s) \quad \dots (4)$$

are subgroups of G .

Now H_{j+1} is normal subgroup of H_j implies that $G_{i,j+1}$ is normal subgroup of $G_{i,j}$. Similarly $H_{k,l+1}$ is normal subgroup of $H_{k,l}$.

Since $H_t = (e)$ and $H_0 = G$, we have

$$G_{i,t} = G_{i+1}^t(G_i \cap H_t) = G_{i+1}^t(e) = G_{i+1}.$$

$$\text{Also } G_{i,0} = G_{i+1}(G_i \cap H_0) = G_{i+1}(G_i \cap G) = G_{i+1} G_i = G_i$$

$$\text{Thus } G_{i,t} = G_{i+1} = G_{i+1,0} \quad \forall i=0, 1, 2, \dots, s-1. \quad \dots (5)$$

$$\text{Similarly } H_{k,s} = H_{k+1} = H_{k+1,0} \quad \forall k=0, 1, 2, \dots, t-1. \quad \dots (6)$$

Consider two series

$$G = G_0 = G_{0,0} \supseteq G_{0,1} \supseteq G_{0,2} \supseteq \dots \supseteq G_{0,t} (= G_1 = G_{1,0}) \supseteq G_{1,1} \supseteq G_{1,2} \supseteq \dots \supseteq G_{1,t} (= G_2 = G_{2,0}) \supseteq \dots \supseteq G_{s-1,0} \supseteq G_{s-1,1} \supseteq \dots \supseteq G_{s-1,t} = G_s = (e) \quad \dots (7)$$

$$G = H_0 = H_{0,0} \supseteq H_{0,1} \supseteq H_{0,2} \supseteq \dots \supseteq H_{0,s} (= H_1 = H_{1,0}) \supseteq H_{1,1} \supseteq H_{1,2} \supseteq \dots \supseteq H_{1,s} (= H_2 = H_{2,0}) \supseteq \dots \supseteq H_{t-1,0} \supseteq H_{t-1,1} \supseteq \dots \supseteq H_{t-1,s} = H_t = (e) \quad \dots (8)$$

Both (7) and (8) have same number of terms, viz, $ts+1$.

Clearly G_0 occurs in (7) and for each $m=1, 2, \dots, s$, as $G_m=G_{m-1}$, by (5), we see that each G_m occurs in (7) for all m . Thus (7) is a refinement of (1). Similarly (8) is a refinement of (2).

Now by Zassenhaus theorem (Theorem 5.21)

$$\frac{G_{rs}}{G_{rs+1}} = \frac{G_{r+1}(G_r \cap H_s)}{G_{r+1}(G_r \cap H_{s+1})} \cong \frac{H_{s+1}(H_s \cap G_r)}{H_{s+1}(H_s \cap G_{r+1})} = \frac{H_{s,r}}{H_{s,r+1}}$$

for all $r=0, 1, 2, \dots, s-1$ and for all $s=0, 1, 2, \dots, t-1$.

Thus (7) and (8) are equivalent. This proves the theorem. ■

Let us come back to composition series now. A group may or may not have composition series. Indeed the following lemma shows that if a commutative group is infinite it cannot have a composition series. In particular the additive group of integers has no composition series.

Lemma 5.23. *If G is a commutative group having a composition series then G is finite.*

Proof: We firstly show that a simple Abelian group must be a cyclic group of prime order. This follows at once from Theorem 3.47(a) and the fact that any subgroup of an Abelian group is normal.

Now let $G=G_0 > G_1 > G_2 > \dots > G_s=(e)$ be a composition series of G . Since $\frac{G_{s-1}}{G_s} = \frac{G_{s-1}}{(e)} \cong G_{s-1}$ is simple Abelian, $o(G_{s-1})=p_{s-1}$

where p_{s-1} is a prime number. Further $\frac{G_{s-2}}{G_{s-1}}$ is simple Abelian $\Rightarrow o\left(\frac{G_{s-2}}{G_{s-1}}\right)=p_{s-2}$, for some prime number p_{s-2} .

We can write G_{s-2} as disjoint union of p_{s-2} distinct right cosets $G_{s-1}a_1, G_{s-1}a_2, \dots, G_{s-1}a_{p_{s-2}}$ and each has p_{s-1} elements. So G_{s-2} has $p_{s-1}p_{s-2}$ elements.

Proceeding in this manner we get that G has $p_0 p_1 p_2 \dots p_{s-1}$ elements where $p_i = o\left(\frac{G_i}{G_{i+1}}\right)$ for $i=0, 1, 2, \dots, s-1$. ■

Now let G be an arbitrary group having a composition series $G=G_0 > G_1 > G_2 > \dots > G_s=(e)$. . . (1)

Suppose $G=H_0 \supseteq H_1 \supseteq H_2 \supseteq \dots \supseteq H_t=(e)$ is a refinement of the former. This refinement can be proper if for some i , H_i is not equal to any G_i . It is clear that for some i , $G_i > H_j > G_{i+1}$. Further we can choose j to be such that $G_i \not\supseteq H_{j-1}$. Then $H_{j-1} \supseteq G_i > H_j > G_{i+1}$. As H_j is normal in H_{j-1} , we get H_j is also normal in G_i . Consequently H_j/G_{i+1} becomes a proper normal subgroup of G_i/G_{i+1} . As G_i/G_{i+1} is simple, this is not possible. Consequently we get that a

composition series can not have proper refinement.

Suppose $G = G_0' > G_1' > \dots > G_u' = (e) \dots (2)$ is another composition series of G . By Schreier's refinement theorem (1) and (2) have equivalent refinements. However, as see earlier none of the two series admit a proper refinement. Therefore, both of them must themselves be equivalent. This establishes the following.

Theorem 5.14. (Jordan-Holder Theorem). *If a group G has a composition series, then all its composition series are pairwise equivalent. ■*

Although we have proved Jordan-Holder theorem for general groups, yet there is another proof for finite groups which is of special interest because of its simplicity and non-dependence on Schreier's refinement theorem. In the following we give that proof.

Theorem 5.25. (Jordan-Holder Theorem for finite groups).

(1) *Every finite group having at least two elements has a composition series.*

(2) if $(e) = G_0 < G_1 < G_2 < \dots < G_k = G$

and $(e) = H_0 < H_1 < H_2 < \dots < H_t = G$

are two composition series for a finite group G then $k=t$ and $\forall i=0, 1, 2, \dots, k-1, \frac{G_{i+1}}{G_i} \cong \frac{H_{\pi(i)+1}}{H_{\pi(i)}}$ for some permutation π of the set $\{0, 1, 2, \dots, k-1\}$ (i.e. any two composition series for G are equivalent).

Proof: (1) We shall apply induction on $o(G)$. Let $o(G) = n$. In case $n=2$, $(e) = G_0 < G_1 = G$ is the only composition series for G as $G_1/G_0 \cong G$ and G being a cyclic group of prime order, is simple. So the theorem holds for $n=2$.

Let every group of order $< n$ have a composition series and let G be a group of order $n > 2$.

If G is simple, then G has no proper normal subgroup. Consequently $(e) = G_0 < G_1 = G$ is the only composition series for G .

If G is not simple let N be a proper normal subgroup of G . Since G is finite, there exist only finitely many proper normal subgroups of G containing N and let M be one such subgroup having largest number of elements. Then M is a maximal normal subgroup of G . Clearly G/M is a simple group and $M \neq G$. So that $o(M) < n$. Hence by the induction hypothesis M has a composition series, say, $(e) = M_0 < M_1 < M_2 < \dots < M_t = M$. Now consider the series

$$(e) = M_0 < M_1 < M_2 < \dots < M_t = M < G. \dots (1)$$

By definition of a composition series, each $\frac{M_{i+1}}{M_i}$ is a simple group for all $i=0, 1, 2, \dots, t-1$. Also $\frac{G}{M_t} = \frac{G}{M}$ is simple. Consequently (1) is a composition series for G .

(2) Suppose

$$(e) = G_0 < G_1 < G_2 < \dots < G_{k-1} < G_k = G \quad \dots (2)$$

$$\text{and} \quad (e) = H_0 < H_1 < H_2 < \dots < H_{l-1} < H_l = G \quad \dots (3)$$

are two composition series for the finite group G . Again we shall apply induction on $o(G)$. Let $o(G) = n$. If $n=2$ then as seen before G has only one composition series, so 2nd part of theorem holds for groups of order 2.

Let us suppose that $o(G) = n > 2$ and let the result be true for all groups of order $< n$. We consider two cases (I) $G_{k-1} = H_{l-1}$ and (II) $G_{k-1} \neq H_{l-1}$.

Case I. $G_{k-1} = H_{l-1}$

$$\text{Evidently } (e) = G_0 < G_1 < G_2 < \dots < G_k \quad \dots (4)$$

$$\text{and} \quad (e) = H_0 < H_1 < H_2 < \dots < H_{l-1} = G_{k-1} \quad \dots (5)$$

are two composition series for the group G_{k-1} whose order is less than n . So by the induction hypothesis series (4) and (5) are equivalent. This gives $k-1 = l-1$ i.e., $k=l$ and further series (2) and (3)

are equivalent as $\frac{G_k}{G_{k-1}} = \frac{G}{G_{k-1}} = \frac{G}{H_{l-1}} = \frac{H_l}{H_{l-1}}$. This case now follows.

Case II. $G_{k-1} \neq H_{l-1}$.

As G_{k-1} and H_{l-1} are maximal normal subgroups of G , $Q = G_{k-1} \cap H_{l-1}$ is normal subgroup of G_{k-1} as well as of H_{l-1} and as $G_{k-1} \neq H_{l-1}$, $Q < G_{k-1}$ and $Q < H_{l-1}$. Further $G_{k-1}H_{l-1}$ is a normal subgroup of G . Now $G_{k-1} \subseteq G_{k-1}H_{l-1}$ and $H_{l-1} \subseteq G_{k-1}H_{l-1}$. Since G_{k-1} and H_{l-1} are maximal normal subgroups of G , $G_{k-1}H_{l-1} = G$ (it can't be equal to G_{k-1} and H_{l-1} simultaneously).

Let $(e) = Q_0 < Q_1 < Q_2 < \dots < Q_m = Q$ be a composition series for the group Q , then we claim that

$$(e) = Q_0 < Q_1 < Q_2 < \dots < Q_m (=Q) < G_{k-1} < G_k = G \quad \dots (6)$$

$$\text{and} \quad (e) = Q_0 < Q_1 < Q_2 < \dots < Q_m (=Q) < H_{l-1} < H_l = G \quad \dots (7)$$

both are composition series for G . For this purpose it is sufficient to prove that $\frac{G_{k-1}}{Q}$ and $\frac{H_{l-1}}{Q}$ are simple.

Now $\frac{G_{k-1}}{Q} = \frac{G_{k-1}}{G_{k-1} \cap H_{l-1}} \cong \frac{G_{k-1}H_{l-1}}{H_{l-1}} = \frac{G}{H_{l-1}}$; but $\frac{G}{H_{l-1}}$ is simple

hence $\frac{G_{k-1}}{Q}$ is simple. Similarly $\frac{H_{l-1}}{Q} \cong \frac{G}{G_{k-1}}$ is simple. Again (6)

and (7) are equivalent, since $\frac{G_{k-1}}{Q} \cong \frac{G}{H_{l-1}}$ and $\frac{G}{G_{k-1}} \cong \frac{H_{l-1}}{Q}$ (as seen above) and each of (6) and (7) is of length $m+2$.

Now by Case (I), series (2) and (6) are equivalent, so $k=m+2$. Again by Case (I), series (3) and (7) are equivalent, therefore $l=m+2$.

Hence $k=l$. Also as series (6) and (7) are equivalent; the series (2) and (3) are equivalent. This completes the proof for the Case (II) ■

PROBLEMS

1. By using the fact that any commutative simple group is of prime order, prove the following:

(a) A finite group G is solvable if and only if it has a composition series, each of whose factor group is cyclic of prime order.

(b) Any solvable group with a composition series is a finite group.

2. Let G be a group having a composition series and H , a normal subgroup of G . Prove that G has a composition series, one of whose terms is H .

[Hint. Let $G = G_0 > G_1 > G_2 > \dots > G_s = (e)$ be a composition series. Consider the normal series $G > H > (e)$. By the Schreier's refinement theorem, they have equivalent refinements.]

3. Let $G \neq (e)$ be a finite group. Prove that if G is solvable then G contains a normal Abelian subgroup $H \neq (e)$; if G is not solvable then G has a normal subgroup $H \neq (e)$ such that $H' = H$.

[Hint. Use the fact that G is solvable if and only if $G^{(n)} = (e)$ for some positive integer n .]

4. Determine a composition series of A_n ($n \neq 4$).

5. Let G be a finite solvable group of order mn such that (m, n)

— 1. Prove that G contains at least one subgroup of order m and any two such subgroups are conjugate.

6. Prove that a finite p -group (p is a prime number) is cyclic if and only if it has only one composition series.

7. Let G be a finite Abelian group of order p^n with $(p, p, \dots, p)$ as its set of invariants. Prove that it has $\frac{(p^n - 1)(p^{n-1} - 1) \dots (p - 1)}{(p - 1)^n}$ composition series.

8. A group G is said to be *nilpotent* if it has a normal series

$$G = G_0 \supseteq G_1 \supseteq G_2 \supseteq \dots \supseteq G_s = (e)$$

such that

$$\frac{G_i}{G_{i+1}} \subseteq Z\left(\frac{G}{G_{i+1}}\right) \quad \forall i = 0, 1, 2, \dots, s-1.$$

Prove the following:

(a) Every nilpotent group is solvable, however converse is not true.

(b) A subgroup and a quotient group of a nilpotent group are nilpotent.

(c) Any commutative group is nilpotent.

(d) If G is a group having a normal subgroup H such that G/H and H both are nilpotent, show by an example that G need not be nilpotent (Compare with Theorem 5.10).

9. Prove that every group of order pq , p and q are prime numbers (not necessarily distinct) is solvable.

[Hint. Use Theorem 5.10.]

10. A group G is said to be *perfect* if $G=G'$.

G is a group such that all its subgroups H ($\neq G$) are solvable, but G is not solvable. Show that G is perfect.

11. Show that if N is a normal subgroup of G and G has a composition series then N has a composition series.

[Hint. Use problem 2 above]

12. Let a group G be direct product of two subgroups H and K . Show that G has a composition series if and only if each of H and K has a composition series.

13. Let $G=H \times K$, then show that G is solvable if and only if each of H and K is solvable.

[REMARK. One part of this problems is Cor. 5.11.]

14. Prove that every group of order p^2q , p, q are primes is solvable.

15. Prove that every group of order pqr ; p, q, r are primes, is solvable.

SURVEY OF SOME FINITE GROUPS

We have seen before, that two finite groups of same order need not be isomorphic e.g. the group S_3 and the cyclic group of order 6, are of same order, but they are not isomorphic. One can ask the following pertinent question. Given a positive integer n , how many non-isomorphic groups of order n exist? Cayley's theorem shows that any group of order n is isomorphic to a subgroup of the symmetric group S_n of degree n . So the above problem is solvable if we can find all the subgroups of S_n of order n and see which of them are mutually non-isomorphic. Since S_n is a finite group (of order $n!$) S_n has only finite number of such subgroups. So one does feel at the outset that the problem is one of verification and is not difficult to solve. But it is very surprising that the problem yet awaits solution. Here we shall answer the problem for some special values of n like p^2 , pq , 8 etc. where p and q are distinct primes. For the case of Abelian finite groups, the problem is completely solved by the use of Fundamental Theorem of Finite Abelian groups as is seen in Theorem 6.3.

Lemma 6.1. Any two finite Abelian groups G and G^1 are isomorphic if and only if their Sylow subgroups are isomorphic.

Proof: Let $o(G) = p_1^{\alpha_1} p_2^{\alpha_2} \dots p_k^{\alpha_k}$ where p_i 's are distinct primes and $\alpha_i > 0$ are integers. Then

$$G = S_1 \oplus S_2 \oplus \dots \oplus S_k \quad \dots (1)$$

where each S_i is the Sylow p_i -subgroup of G .

If σ is an isomorphism of G onto G^1 then

$$G^1 = \sigma(G) = \sigma(S_1) \oplus \sigma(S_2) \oplus \dots \oplus \sigma(S_k).$$

$$\text{Consequently } o(G^1) = o[\sigma(G)] = \prod_{i=1}^k o[\sigma(S_i)] = \prod_{i=1}^k p_i^{\alpha_i} = o(G)$$

and $o(\sigma(S_i)) = o(S_i) = p_i^{\alpha_i}$ gives that $\sigma(S_1), \sigma(S_2), \dots, \sigma(S_k)$ are the Sylow subgroups of G^1 . Now σ maps S_i isomorphically onto $\sigma(S_i)$. So that $S_i \cong \sigma(S_i) \forall i = 1, 2, 3, \dots, k$. This proves that the condi-

tion is necessary.

Conversely let G and G^1 have isomorphic Sylow subgroups. Since $S_1, S_2, \dots, S_k$ are the only Sylow subgroups of G , G^1 has corresponding Sylow subgroups $S_1', S_2', \dots, S_k'$ such that $S_i \cong S_i' \forall i$ under an isomorphism σ_i of S_i onto S_i' ,

$$\text{Now } G^1 = S_1' \oplus S_2' \oplus \dots \oplus S_k' \quad \dots (2)$$

Define $\sigma : G \rightarrow G^1$ as under.

Take $x \in G$, (1) yields $x = x_1 + x_2 + \dots + x_k$ for some unique $x_i \in S_i \forall i$. Then put $\sigma(x) = \sigma_1(x_1) + \sigma_2(x_2) + \dots + \sigma_k(x_k)$. Clearly $\sigma(x) \in G^1$. For any $x = x_1 + x_2 + \dots + x_k$, $y = y_1 + y_2 + \dots + y_k \in G$ with $x_i, y_i \in S_i \forall i$, we have

$$x + y = (x_1 + y_1) + (x_2 + y_2) + \dots + (x_k + y_k).$$

$$\begin{aligned} \text{Therefore, } \sigma(x + y) &= \sigma_1(x_1 + y_1) + \sigma_2(x_2 + y_2) + \dots + \sigma_k(x_k + y_k) \\ &= [\sigma_1(x_1) + \sigma_1(y_1)] + [\sigma_2(x_2) + \sigma_2(y_2)] + \dots + [\sigma_k(x_k) + \sigma_k(y_k)] \\ &= [\sigma_1(x_1) + \sigma_2(x_2) + \dots + \sigma_k(x_k)] + [\sigma_1(y_1) + \sigma_2(y_2) + \dots + \sigma_k(y_k)] \\ &= \sigma(x) + \sigma(y). \end{aligned}$$

This proves that σ is a homomorphism.

Now

$$\begin{aligned} \sigma(x) = \sigma(y) &\Rightarrow \sigma_1(x_1) + \sigma_2(x_2) + \dots + \sigma_k(x_k) = \sigma_1(y_1) + \sigma_2(y_2) + \dots + \sigma_k(y_k) \\ &\Rightarrow \sigma_i(x_i) = \sigma_i(y_i) \forall i = 1, 2, \dots, k \text{ because of (2)} \\ &\Rightarrow x_i = y_i \forall i = 1, 2, \dots, k; \text{ since } \sigma_i \text{ is } 1-1. \\ &\Rightarrow x = y. \end{aligned}$$

Thus σ is also 1-1.

Consider any $x' \in G^1$; (2) yields, $x' = x_1' + x_2' + \dots + x_k'$ where $x_i' \in S_i' \forall i$. Since $\sigma_i : S_i \rightarrow S_i'$ is onto, there exists $x_i \in S_i$ such that $\sigma_i(x_i) = x_i'$. Take $x = x_1 + x_2 + \dots + x_k$; then

$$\begin{aligned} \sigma(x) &= \sigma_1(x_1) + \sigma_2(x_2) + \dots + \sigma_k(x_k) \\ &= x_1' + x_2' + \dots + x_k' = x'. \end{aligned}$$

Hence σ is onto. Consequently $\sigma : G \rightarrow G^1$ is an isomorphism and $G \cong G^1$. This establishes the lemma. ■

In a worked-out exercise given after section 5, chapter 4 we defined 'partition' of a natural number n . We re-state (slight differently) that definition as under.

Definition 6.2. (Partition of a natural number).

Let n be any natural number. Then any expression of n as a sum of natural numbers is called a *partition* of n . The number of partitions of n will be denoted by $t(n)$.

REMARK. In the expression of n as a sum of natural numbers we write the numbers in descending order of magnitude.

Thus if $n = n_1 + n_2 + \dots + n_k$, $n_1 \geq n_2 \geq \dots \geq n_k$.

EXAMPLE 1. $4 = 4$, $4 = 3 + 1$, $4 = 2 + 1 + 1$, $4 = 2 + 2$, $4 = 1 + 1 + 1 + 1$ are five different partitions of 4. Thus $t(4) = 5$.

EXAMPLE 2 $5=5$, $5=4+1$, $5=3+1+1$, $5=3+2$, $5=2+2+1$, $5=2+1+1+1$, $5=1+1+1+1+1$ are the seven partitions of 5 and so $t(5)=7$.

Theorem 6.3. For any positive integer $n=p_1^{\alpha_1} p_2^{\alpha_2} \dots p_k^{\alpha_k}$ where p_i 's are distinct prime numbers and $\alpha_i > 0$ are integers, the number of non-isomorphic Abelian groups of order n is the product $t(\alpha_1)t(\alpha_2) \dots t(\alpha_k)$.

Proof: Let G be an Abelian group of order n . Then

$$G = S_1 \oplus S_2 \oplus \dots \oplus S_k \quad \dots (1)$$

where each S_i is the Sylow p -subgroup of G of order $p_i^{\alpha_i}$. By Lemma 6.1 any two finite Abelian groups of order n are isomorphic if and only if their corresponding Sylow p_i -subgroups are isomorphic. Thus if λ_i is the number of non-isomorphic Abelian groups of order $p_i^{\alpha_i}$, then there exist $\lambda_1 \lambda_2 \dots \lambda_k$ number of non-isomorphic Abelian groups of order n . So if we can establish that $\lambda_i = t(\alpha_i)$ the result will follow.

Consider any prime p and any positive integer α . We are required to prove that the number of non isomorphic Abelian groups of order p^α is $t(\alpha)$. Consider any Abelian group S of order p^α . Then

$$S = A_1 \oplus A_2 \oplus \dots \oplus A_l \quad \dots (2)$$

where A_i 's are cyclic p -subgroups of order p^{β_i} such that $\beta_1 \geq \beta_2 \geq \dots \geq \beta_l$.

$$(2) \Rightarrow p^\alpha = o(S) = o(A_1)o(A_2) \dots o(A_l) = p^{\beta_1 + \beta_2 + \dots + \beta_l}.$$

Thus $\alpha = \beta_1 + \beta_2 + \dots + \beta_l$ is a partition of α . We know that $p^{\beta_1}, p^{\beta_2}, \dots, p^{\beta_l}$ are the invariants and any two finite Abelian p groups are isomorphic if and only if they have same sets of invariants. Consequently two isomorphic finite Abelian p groups of order p^α determine uniquely same partition of α . Consider a partition $\alpha = r_1 + r_2 + \dots + r_m$ of α . We can take $r_1 \geq r_2 \geq \dots \geq r_m$ by renumbering, if necessary. Choose cyclic groups B_i of order p^{r_i} ($i=1, 2, \dots, m$). Then $S = B_1 \oplus B_2 \oplus \dots \oplus B_m$ is an Abelian p -group of order p^α such that $p^{r_1}, p^{r_2}, \dots, p^{r_m}$ are its invariants. So we have seen that two Abelian p -groups of order p^α are isomorphic if and only if they determine same partition of α . Further to each partition of α there corresponds an Abelian group of order p^α such that it determines the given partition of α . This proves that the number of non-isomorphic Abelian groups of order p^α is same as $t(\alpha)$. Hence the theorem follows. ■

Theorem 6.4. For any prime number p , a group of order p^2 is either cyclic or a direct sum of two cyclic groups of order p each.

Proof: Let G be a group of order p^2 . We know that G is Abelian (Corollary 3.57). If G has an element of order p^2 then G is cyclic.

Suppose that G is not cyclic. Then for each $a(a \neq e) \in G$, $o(a) \mid p$ gives $o(a) = p$. Choose $a(a \neq e) \in G$ then $H = \langle a \rangle$ is a cyclic subgroup of G of order p . Choose $b \in G$ such that $b \notin H$. Then $K = \langle b \rangle$ is also of order p . Now $H \cap K = (e)$. Consequently $o(HK) = o(H)o(K) = p^2 = o(G)$, so that $G = HK$. Hence $G = H \times K$, a direct product of two cyclic groups of order p each. ■

REMARK. The above result follows immediately from Theorem 6.3. Since $2 = 2$, $2 = 1 + 1$ are the only partitions of 2, there are two non-isomorphic Abelian groups of order p^2 having (p^2) and (p, p) as their respective invariants. In the former case G is a cyclic group of order p^2 and in the latter case G is a direct product of two cyclic groups, each of order p .

Theorem 6.5. Let p and q be two distinct prime numbers. To be definite let $p < q$. If G is a group of order pq , then one of the following holds.

(I) G is cyclic (Abelian).

(II) It has two generators a and b such that $a^p = e$, $b^q = e$, $a^{-1}ba = b^r$ for some r such that $r \not\equiv 1 \pmod{q}$, $r^p \equiv 1 \pmod{q}$ and p divides $q-1$ (Non-Abelian).

Proof: Since $o(G) = pq$, G has at least one Sylow q -subgroup of order q . By Sylow's Third Theorem the number of Sylow q -subgroups is $1 + kq$ for some $k \geq 0$ and $1 + kq$ divides p . As $q > p$ we have $k = 0$. Thus G has only one Sylow q -subgroup, which is normal and being of prime order q , is cyclic, say $\langle b \rangle$. Again G has at least one Sylow p -subgroup of order p and number of such Sylow p -subgroups of G is $1 + kp$ for some $k \geq 0$ and $1 + kp$ divides q . As q is prime we have $1 + kp = 1$ or $1 + kp = q$. In the former case G has only one Sylow p -subgroup $\langle a \rangle$; which is normal. Then $\langle a \rangle \cap \langle b \rangle = (e)$ gives $ab = ba$ [since if H and K are two normal subgroups of a group G such that $H \cap K = (e)$, then $hk = kh \forall h \in H, k \in K$] Consequently $o(ab) = o(a)o(b) = pq = o(G)$. Hence $G = \langle ab \rangle$ is a cyclic group of order pq .

We are left with the case when $1 + kp = q$. This implies p divides $q-1$. As G has at least one subgroup, say $\langle a \rangle$ of order p , $o(a) = p$. Now $a^{-1}ba \in \langle b \rangle$, as $\langle b \rangle$, is normal subgroup of G . Consequently $a^{-1}ba = b^r$ for some r with $1 \leq r \leq q$

If $r = 1$ then again we get $G = \langle ab \rangle$, a cyclic group of order pq . So if G is non-Abelian $r \neq 1$. Thus $r \not\equiv 1 \pmod{q}$. Now $a^{-1}ba^2 = a^{-1}(a^{-1}ba)a = a^{-1}b^ra = (a^{-1}ba)^r = (b^r)^r = b^{r^2}$. Similarly $a^{-3}ba^3 = b^{r^3}$ and so on, finally $a^{-p}ba^p = b^{r^p}$. However $a^p = e$. Hence $b = b^{r^p}$. This yields $r^p \equiv 1 \pmod{q}$. Further notice that $\langle a \rangle \cap \langle b \rangle = (e)$ and hence $o[\langle a \rangle \langle b \rangle] = o\langle a \rangle o\langle b \rangle = pq = o(G)$. Hence $G = \langle a \rangle \langle b \rangle$. Thus each element of G is of the form $a^u b^v$. This proves that G is

a group generated by two elements a and b such that $a^p = e$, $b^q = e$, $a^{-1}ba = b^r$, $r \not\equiv 1 \pmod{q}$, $r^p \equiv 1 \pmod{q}$ and $p \mid (q-1)$.

Conversely let r, p, q be such that $p < q$ are primes and they satisfy the above conditions. Consider two cyclic groups $H = \langle a \rangle$ and $K = \langle b \rangle$ of orders p and q respectively. Let G be the set $H \times K = \{(a^u, b^v) \mid u \text{ and } v \text{ are integers such that } 1 \leq u \leq p-1 \text{ and } 1 \leq v \leq q-1\}$. G has pq elements. For any $(a^u, b^v), (a^x, b^y)$ in G , define $(a^u, b^v) \cdot (a^x, b^y) = (a^{u+x}, b^{vr^x+y})$. . . (1)

To show that the composition is well-defined, let $(a^u, b^v) = (a^{u_1}, b^{v_1}), (a^x, b^y) = (a^{x_1}, b^{y_1})$

$$\text{This } \Rightarrow a^u = a^{u_1}, b^v = b^{v_1}, a^x = a^{x_1}, b^y = b^{y_1}$$

$$\Rightarrow p \mid (u - u_1), p \mid (x - x_1), q \mid (v - v_1) \text{ and } q \mid (y - y_1) \quad \dots (2)$$

$$\Rightarrow p \mid (u+x) - (u_1+x_1) \Rightarrow a^{u+x} = a^{u_1+x_1}, \text{ since } o(a) = p. \quad \dots (3)$$

$$\text{Again consider } vr^x + y - v_1r^{x_1} - y_1 \\ = v(r^x - r^{x_1}) + r^{x_1}(v - v_1) + y - y_1. \quad \dots (4)$$

$$\text{Since } q \mid (v - v_1) \text{ and } q \mid (y - y_1), \text{ if } x = x_1, \text{ we get} \quad \dots (4)$$

from (4) that $q \mid (vr^x + y) - (v_1r^{x_1} + y_1) \Rightarrow b^{vr^x+y} = b^{v_1r^{x_1}+y_1}$.

Suppose $x \neq x_1$. To be definite let $x < x_1$. Then

$$r^x - r^{x_1} = r^x(1 - r^{x_1-x}). \text{ As } r^p \equiv 1 \pmod{q} \text{ and } p \mid (x_1 - x) \text{ we have} \\ r^{x_1-x} \equiv 1 \pmod{q} \Rightarrow q \mid (1 - r^{x_1-x}) \Rightarrow q \mid (r^x - r^{x_1}).$$

Then again, (4) yields that $b^{vr^x+y} = b^{v_1r^{x_1}+y_1}$.

Hence $(a^{u+x}, b^{vr^x+y}) = (a^{u_1+x_1}, b^{v_1r^{x_1}+y_1})$.

This proves the composition is well defined

Consider $(a^u, b^v), (a^s, b^y), (a^s, b^t) \in G$.

$$\begin{aligned} (1) \text{ Associativity: } & [(a^u, b^v)(a^s, b^y)](a^s, b^t) \\ & = (a^{u+s}, b^{vr^s+y})(a^s, b^t) \\ & = (a^{u+s+s}, b^{(vr^s+y)r^s+t}) \\ & = (a^{u+s+s}, b^{vr^{s+s}+yr^s+t}) \end{aligned}$$

$$\text{Similarly } (a^u, b^v)[(a^s, b^y)(a^s, b^t)] \\ = (a^{u+s+s}, b^{vr^{s+s}+yr^s+t}).$$

This proves the law of associativity.

$$(2) \text{ Existence of Identity: } e = (a^0, b^0) \text{ is such that} \\ (a^0, b^0)(a^x, b^y) = (a^{0+x}, b^{0r^x+y}) \\ = (a^x, b^y) = (a^x, b^y)(a^0, b^0).$$

Thus (a^0, b^0) is the identity of G .

(3) Existence of right inverse:

Taking $x = p - u$ and y to be such that $vr^x + y$ is divisible by q ; for example take $y = lq - vr^x$ where l is a positive integer such that $lq > vr^x$.

$$\text{Then } (a^u, b^v)(a^x, b^y) = (a^u, b^{vr^x+y}) = (a^0, b^0).$$

Hence G is a group.

Note let $a_1 = (a, b^0)$, $b_1 = (a^0, b)$ then

$a_1^u = (a^u, b^0)$, $b_1^v = (a^0, b^v)$; so $a_1^u b_1^v = (a^u, b^v)$ and therefore, a_1 and b_1 generate G . Further $a_1^p = (a^p, b^0) = (a^0, b^0) = e$, the identity of G . Similarly $b_1^q = e$. Now $a_1^{-1} b_1 a_1 = (a^{p-1}, b^0)(a^0, b)(a, b^0) = (a^0, b^r) = b_1^r$. Hence G is a group of order pq generated by elements a_1 , b_1 of orders p and q respectively, $r \not\equiv 1 \pmod{q}$, $r^p \equiv 1 \pmod{q}$ and $p \mid (q-1)$.

Now we show that any two non-Abelian groups G and G' of order pq are isomorphic.

Now G and G' have generating sets $\{a, b\}$ and $\{a', b'\}$ such that $a^p = e$, $b^q = e$, $(a')^p = e'$, $(b')^q = e'$, for some integers r, r' ; $a^{-1}ba = b^r$, $(a')^{-1}b'a' = (b')^{r'}$ with $r \not\equiv 1 \pmod{q}$, $r' \not\equiv 1 \pmod{q}$, $r^p \equiv 1 \pmod{q}$, $(r')^p \equiv 1 \pmod{q}$. Now all solutions of the congruence $x^p \equiv 1 \pmod{q}$ which are not congruent to 1 modulo q are congruent to some one of $r, r^2, r^3, \dots, r^{p-1}$ modulo q . Thus we have $r' \equiv r^i \pmod{q}$ for some i , between 1 and $p-1$. Then $b^{r'} = b^{r^i} = a^{-i}ba^i$. If we put $a_1 = a^i$, $b_1 = b$ we see that a_1, b_1 also generate G and $a_1^{-1} b_1 a_1 = b_1^{r'}$. Define a mapping $\sigma: G \rightarrow G'$ by $\sigma(a_1^u b_1^v) = (a')^u (b')^v \forall$ integers u, v . We leave it to the readers to verify that σ is a well-defined mapping and is an isomorphism. ■

REMARK. The above theorem gives that for any two prime numbers $p < q$. Following cases happen.

(1) If p does not divide $q-1$, then any group G of order pq is cyclic.

(2) If p divides $q-1$ then there are only two non-isomorphic groups of order pq , one of which is commutative (which is again cyclic as p and q are different primes) other is non-commutative.

EXAMPLE 3. Consider $p=2$, $q=5$. Clearly $p \mid (q-1)$. Thus there are two non-isomorphic groups of order 10 one of them is cyclic and other is generated by two elements a and b such that $a^2=e$, $b^5=e$, $a^{-1}ba=b^4$, since $4^2 \equiv 1 \pmod{5}$.

EXAMPLE 4. Consider $p=3$, $q=5$. Here p does not divide $q-1$. Thus any group of order 15 is cyclic.

Definition 6.6. For any positive integer n , a group D_n generated by two elements a and b satisfying the following relations $a^n=e$, $b^2=e$, $bab^{-1}=a^{-1}$; is called n th dihedral group.

Let $H = \langle a \rangle$ and $K = \langle b \rangle$ be cyclic groups of orders n and 2 respectively. Consider the set $G = H \times K = \{(a^u, b^v) \mid u, v \text{ any integers}\}$. If, for any $(a^u, b^v), (a^x, b^y) \in G$ we define $(a^u, b^v)(a^x, b^y) = (a^{u+x}, b^y)$ in case $v=0$ and $(a^u, b^v)(a^x, b^y) = (a^{u-x}, b^{v+y})$ in case $v=1$. Then G is a group of order $2n$ with $a_1 = (a, b^0)$, $b_1 = (a^0, b)$, as its generators. Also $a_1^n = e$, $b_1^2 = e$, $b_1 a_1 b_1^{-1} = a_1^{-1}$. Hence G is an n th

dihedral group. It can be verified that any n th dihedral group is isomorphic to G .

Definition 6.7. A group G generated by two elements a and b satisfying the relations $a^4=e$, $b^2=a^2$, $bab^{-1}=a^{-1}$ is called *group of Quaternions*.

The set H consisting of the following matrices

$$\begin{aligned} &\pm \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \pm \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \pm \begin{pmatrix} \sqrt{-1} & 0 \\ 0 & -\sqrt{-1} \end{pmatrix}, \\ &\pm \begin{pmatrix} 0 & \sqrt{-1} \\ \sqrt{-1} & 0 \end{pmatrix} \end{aligned}$$

is a group of order 8 under usual matrix multiplication. If we take

$$a = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \text{ and } b = \begin{pmatrix} \sqrt{-1} & 0 \\ 0 & -\sqrt{-1} \end{pmatrix},$$

it can be easily seen that a and b satisfy the conditions defining the quaternion group. Any group of quaternions can be seen to be isomorphic to H .

Theorem 6.8. Any non-Abelian group G of order 8 is either D_4 or the group of quaternions.

Proof: Since G is not Abelian, G is not cyclic so G cannot have any element of order 8. Further if every non-identity element of G is of order 2, G will again be Abelian. Thus there exists a $(\neq e) \in G$ such that $o(a)$ is neither 2 nor 8. Then $o(a) \mid o(G)$ gives $o(a)=4$. Thus the subgroup $H=\{e, a, a^2, a^3\}$ being of index 2 is normal in G . Choose $b \in G$ such that $b \notin H$.

Then $G=H \cup bH=\{e, a, a^2, a^3, b, ba, ba^2, ba^3\}$.

This shows that a and b generate G . If $ab=ba$ then G will be commutative. So $ab \neq ba$. Now $bab^{-1} \in H$, as H is normal, $o(bab^{-1})=o(a)=4$. In H , only element other than a of order 4 is a^3 . So that $bab^{-1}=a^3=a^{-1}$. . . (1)

Now G/H is of order 2, so $(bH)^2=H \Rightarrow b^2 \in H$. This yields $b^2=e$ or $b^2=a$ or $b^2=a^2$ or $b^2=a^3$.

If $b^2=a$ or $b^2=a^3$ then $o(b)=8$. This is not possible as G is not commutative. Hence $b^2=e$ or $b^2=a^2$. In the former case we get that G is D_4 and in the latter case we get that G is the group of quaternions. ■

REMARK. Since $3=3$, $3=2+1$, $3=1+1+1$ are the three partitions of 3, there are three non-isomorphic Abelian groups of order 8. Above theorem gives that there are two non-isomorphic non-Abelian groups of order 8. Hence there are five non-isomorphic groups of order 8.

PROBLEMS

1. By taking $p=2$ in Theorem 6.5, prove that the congruence relation $r^2 \equiv 1 \pmod{q}$ has only two solutions viz. $r \equiv 1 \pmod{q}$, $r \equiv -1 \pmod{q}$. Conclude that any group of order $2q$ (q is an odd prime) is either cyclic or the dihedral group D_q .
2. Prove that any non-commutative group of order 6 is isomorphic to the symmetric group S_3 .
3. Prove that in S_n , the sum of lengths of any set of non-identity pairwise disjoint cycles cannot exceed n . Using the fact that any group of order 15 is cyclic, conclude that A_5 , A_6 and A_7 have no subgroups of order 15, although 15 divides their orders.
4. If a group G of order 12 is not isomorphic to A_4 , prove that G must contain an element of order 6.
5. If a non-Abelian group of order 12 is not isomorphic to A_4 and contains at least two elements of order 2, then G is the dihedral group D_6 .
6. Show the direct product $S_3 \times H$, where H is a cyclic group of order 2, is isomorphic to D_6 .
7. Show that the number of non-isomorphic Abelian groups of order 12 is two.

MISCELLANEOUS WORKED-OUT EXERCISES

Exercise 1. Let f, g be two disjoint cycles in S_n of order m and k respectively. Show that $o(fg) = d$ where $d = [m, k]$.

Solution. $o(f) = m$, $o(g) = k \Rightarrow f^m = I$ and $g^k = I$. Since f and g are disjoint, $fg = gf$.

So $(fg)^d = f^d g^d = I$, as $m \mid d$ and $k \mid d$. Further let $(fg)^r = e$ then $f^r = g^{-r}$.

If f^r is not the identity permutation, there exists $a \in S$ such that $f^r(a) \neq a$, so $f^r(a)$ belongs to the cycle of f . But $f^r(a) = g^{-r}(a)$ implies $g^{-r}(a)$ belongs to cycle of f . This is absurd as f and g are disjoint cycles. Hence our assumption is wrong. Thus $f^r = I = g^{-r} \Rightarrow f^r = I$ and $g^r = I \Rightarrow m \mid r$ and $k \mid r$. Now $d = [m, k]$. Consequently $d \mid r$. Hence $o(fg) = d$.

Exercise 2. Show that a group of order 112 can not be simple.

Solution. Let G be a group of order 112. $o(G) = 112 = 2^4 \cdot 7$.

Consider Sylow 2-subgroups. Number of Sylow 2-subgroups is $1 + 2k$ where $k \geq 0$ is an integer, such that $(1 + 2k) \mid 7$. This gives $k = 0, 3$.

In case $k = 0$, we have only one subgroup of order 16. So it must

be normal and hence G can not be a simple group.

In case $k=3$, we have 7 Sylow 2-subgroups of G , and all of them are conjugate to each other. Let H be one of these Sylow 2-subgroups. Number of distinct right cosets of H is $o(G)/o(H)=7$; let these be $Hx_1, Hx_2, \dots, Hx_7$.

Define $f: G \rightarrow S_7$ by $f(x)=T_x \forall x \in G$ where $T_x(Hx_i)=Hx_i x^{-1} \forall i=1, 2, \dots, 7$. It can be easily checked that $T_x \in S_7$, f is a homomorphism, and $\text{Ker } f \subseteq H$.

If G were a simple group, $\text{Ker } f = \{e\}$, where e is identity of G and $G \neq \text{Ker } f$ otherwise $G=H$!

Now since $\text{Ker } f = \{e\}$ we get that $G \cong f(G)$, where $f(G)$ is a subgroup of S_7 .

If $f(G)$ consists of all even permutations, then $f(G) \subseteq A_7$. But $o(A_7)=2520$ and $o[f(G)]=o(G)=112 \Rightarrow 112 \nmid 2520$. This is not true. Thus $f(G)$ has at least one odd permutation. The set of even permutations in $f(G)$ form a normal subgroup of $f(G)$ of index 2, which is against the simplicity of G . Consequently G can not be a simple group.

Exercise 3. Characterize all groups G which have exactly one proper subgroup.

Solution. By problem 24, section 3, Chapter 2, G must be a finite group. Let p and q be two distinct prime numbers dividing $o(G)$. Then by Sylow's Theorem, G has subgroups of order p and q . This is against the hypothesis of the exercise. Hence $o(G)=p^n$ where p is a prime number and n is an integer ≥ 2 .

Again by Sylow's theorem, G has subgroups of order p^{n-1} and p . Since G has exactly one proper subgroup, $n-1=1$, i.e., $n=2$. Consequently G is an Abelian group of order p^2 . Either G is a cyclic group or G is the direct product of two cyclic groups each of order p . In the second case G has more than one proper subgroup of order p , which is once again, a contradiction to the hypothesis. Hence G is a cyclic group of order p^2 where p is a prime number.

Exercise. Definition: A subgroup H of a group G is called a maximal subgroup if (i) $H \neq G$, and (ii) There does not exist any subgroup K of G , such that $H < K < G$.

Show that $\langle \mathbb{Q}, + \rangle$ has no maximal subgroup.

Solution. Let H be a maximal subgroup of $\mathbb{Q}$. There exists a rational number r/s ($r, s \in \mathbb{Z}$, $s \neq 0$) such that $r/s \notin H$. As $\mathbb{Z}$ is a subgroup of $\mathbb{Q}$, and $\mathbb{Z} \neq \mathbb{Q}$; $H \neq \{0\}$. Let $m/n \in H$, $m \in \mathbb{Z}$, $n \in \mathbb{Z}$, $mn \neq 0$.

Now $m/n \in H \Rightarrow n.m/n \in H \Rightarrow m \in H$. Maximality of H implies $H + \langle r/s \rangle = \mathbb{Q}$. As a consequence $r/sms = h + t r/s$ for some $h \in H$,

$t \in \mathbb{Z}$. Thus $r/s = msh + tmr = (sh)m + (tr)m \in H$ as $m \in H$, $sh, tr \in \mathbb{Z}$. This is a contradiction.

Exercise 5. Prove that the centre of a group G is properly contained in every maximal subgroup of G having a composite index.

Solution. Let H be a maximal subgroup of G , with composite index.

Let Z be the centre of G and let Z be not contained in H . There exists $a \in Z$, $a \notin H$.

Consider $\langle H, a \rangle$ i.e., the subgroup generated by H and a . Since H is maximal; $H \subseteq \langle H, a \rangle$ and $a \notin H \Rightarrow \langle H, a \rangle = G$. So any $g \in G$ can be written as $a^i h$ for some $h \in H$, $i \in \mathbb{Z}$. (Note that we have used the fact that $a \in Z$).

Now for any $y \in H$, $g \in G$, $gyg^{-1} = (a^i h)y(a^i h)^{-1} = a^i h y h^{-1} a^{-i} = h y h^{-1}$ as $a \in Z$. But $h, y, h^{-1} \in H$ imply that $gyg^{-1} \in H$. Hence H is a normal subgroup of G . As G/H has composite order, it has a non-trivial subgroup say K/H , i.e., $\{H\} < K/H < G/H$. This in turn gives that $H < K < G$. This is against the maximality of G . Hence $Z \subseteq H$.

Now if $Z=H$ then H is a normal subgroup of G and by the same arguments as given above, we get a contradiction. Hence $Z < H$.

Exercise 6. If G is a perfect group, then centre of $G/Z(G)$ is trivial.

Solution. For convenience we shall write Z for $Z(G)$. Suppose centre of $G/Z(G)$ is non-trivial. There exists $Za \in Z\left(\frac{G}{Z}\right)$, such that $a \notin Z$. Define a mapping $f: G \rightarrow G$ by $f(x) = axa^{-1}x^{-1}$. We claim that $axa^{-1}x^{-1} \in Z$ for all $x \in G$. Since $Za \in Z\left(\frac{G}{Z}\right)$, $ZaZx = ZxZa \Rightarrow Zax = Zxa \Rightarrow (ax)(xa)^{-1} \in Z \Rightarrow axa^{-1}x^{-1} \in Z$.

Thus f is a mapping from G to Z . We now show that f is a homomorphism.

$$\begin{aligned} \text{Note that } f(xy) &= axya^{-1}y^{-1}x^{-1} = (axa^{-1}x^{-1})x(aya^{-1}y^{-1})x^{-1} \\ &= (axa^{-1}x^{-1})(aya^{-1}y^{-1})xx^{-1}, \text{ as } (aya^{-1}y^{-1}) \in Z \\ &= f(x)f(y). \end{aligned}$$

Now $\text{Ker } f$ is a normal subgroup of G .

If $\text{Ker } f = G$ then for all $x \in G$, $f(x) = e \Rightarrow axa^{-1}x^{-1} = e \Rightarrow ax = xa \Rightarrow a \in Z$ which is a contradiction. Hence $\text{Ker } f < G$.

Finally as $f: G \rightarrow Z$ is a homomorphism, $G/\text{Ker } f \cong T$ where T is a subgroup of Z . As Z is Abelian, $G/\text{Ker } f$ is Abelian hence $G' \subseteq \text{Ker } f$.

But G is perfect, $G=G'$. This in turn implies $G = \text{Ker } f$ which is absurd. Hence $aZ = Z$; as a consequence, centre of G/Z consists of Z only.

Exercise 7. (Wilson's Theorem). If p is a prime number then $(p-1)! \equiv -1 \pmod{p}$.

Solution. Consider the symmetric group S_p .

As $p \mid p!$, $p^2 \nmid p!$ Sylow p -subgroup of S_p is of order p . Number of Sylow p -subgroups of S_p is $1+kp$. As each Sylow p -subgroup is of order p , no two Sylow p -subgroups have any element in common except identity. Then the total number of elements of order p is $(1+kp)(p-1)$.

The elements of order p in S_p are precisely the cycles of length p . So their number is $\left(\frac{1}{p}\right)p!/0! = (p-1)!$. Hence $(p-1)! = (1+kp)(p-1) = p + kp^2 - kp - 1 \Rightarrow (p-1)! \equiv -1 \pmod{p}$.

MISCELLANEOUS PROBLEMS

- For any non-empty subset H of G , define $H^{-1} = \{x^{-1} \mid x \in H\}$. Prove : (i) H is a subgroup of $G \Rightarrow H^{-1} = H$. But converse of this result need not be true in general.
(ii) If H is a non-empty set then H is a subgroup of G if and only if $HH^{-1} \subseteq H$.
- Show that a subgroup H of a group G is normal subgroup of G if and only if for all $x, y \in G$ $xy \in H \Rightarrow yx \in H$.
- Give an example of a non-Abelian group G such that $(xy)^3 = x^3y^3$ for all x, y , in G .
[Hint. Consider $S = \{(x, y, z) \mid x, y, z \in \mathbb{Z}/\langle 3 \rangle\}$. Define $*$ on S by $(x, y, z) * (x', y', z') = (x+x', y+y', z+z'+2xy')$.]
- Let G be a finite Abelian group. Prove that the number of solutions of $x^n = e$ where $n \mid \phi(G)$ and e is identity of G , is a multiple of n .
- Show that A_6 has no subgroup of order 120
[Hint. Use simplicity of A_6 .]
- Show that a group of order 120 is not simple.
[Hint. Use problem 5 above.]
- Prove that a group of order $p \cdot q$, p and q being distinct prime numbers, is not simple. Show further that if $q < p$ and $q \nmid (p^2-1)$ then every group of order $p \cdot q$ is Abelian.
- Prove that $xax = b$ is solvable for x in G if and only if ab is the square of some element in G .
- Characterize all groups, every non-identity element a of which commutes with a, a^{-1} and e only.
[Ans. Cyclic groups of order 1, 2, 3 and S_3]

10. Let G be a group of order $2n$. Suppose that half of the elements are of order 2 and the other half form a subgroup H of order n . Prove that H is of odd order and is an Abelian subgroup of G .

Give an example of a group satisfying the conditions given above.

[Hint. Try S_3 .]

11. Prove that every finite group having more than two elements has a non-trivial automorphism.

12. Give an example of an *outer* automorphism of a group G (i.e. an automorphism of G which is not inner) which maps each conjugate class of G onto itself.

[Hint. Take $G = \left\{ \begin{pmatrix} 1 & a \\ 0 & b \end{pmatrix} \mid a, b \in \mathbb{Z}_8 \text{ and } (b, 8) = 1 \right\}$. Show that G is group under matrix multiplication. Define $f: G \rightarrow G$ by

$$f\left(\begin{pmatrix} 1 & a \\ 0 & b \end{pmatrix}\right) = \begin{pmatrix} 1 & a + \frac{b^2 - 1}{2} \\ 0 & b \end{pmatrix}.$$

13. Prove that a finite group of order $2n$ which contains an element of order n has at least $\tau(n) + 1$ normal subgroups, where $\tau(n)$ is the number of divisors of n .

14. Let G be an additive Abelian group. For any non-void subset S of G , let $D(S)$ denote the set of differences $x - y$, where $x, y \in S$. Show that if A and B are any non-void subsets of G such that $G = A \cup B$, then either $D(A) \supseteq B$ or $D(B) \supseteq A$. Show that if $G = A \cup B$ and if A and B are not disjoint then $D(A) = G$ or $D(B) = G$.

15. Let S be the set of all elements of G which are of order n , a fixed positive integer. Prove that the subgroup generated by S is a normal subgroup of G .

16. Let G be a finite group, show that G is a cyclic p -group if and only if G has a unique maximal subgroup.

17. Prove that the n th roots of unity constitute the *only* set of n th-roots of a complex number which is a group under multiplication.

[Hint. Let z be a complex number whose n th-roots form a group G under multiplication. Clearly G is a subgroup of non-zero complex numbers under multiplication. So $1 \in G \Rightarrow 1 = z^{1/n} \Rightarrow z = 1$.]

18. Let G be a finite group and suppose that for all $d \geq 1$, G has at most d elements of order d . Prove that G is cyclic.

19. Let G be a finite group of odd order. Prove that the set of products of all elements of G taken in any order, is contained in the commutator subgroup.

20. G is a finite group. Show that every element of G is expres-

sible as a square of some element of G if and only if $o(G)$ is odd.

21. If p is a prime dividing n then prove that the only elements of order p in S_n are products of disjoint p -cycles.

[Hint. Use worked out Exercise 1 after Theorem 6.8.]

22. Show that A_5 has no subgroup of order 15 and 20.

[Hint. Every group of order 15 is cyclic and A_5 has no element of order 15.]

23. Let S be a Sylow 2-subgroup of A_5 . Show that S has 5 conjugates in A_5 .

RINGS

1. Introduction

Up-till now we have considered sets with one binary composition only. But there are non-void sets with more than one binary compositions namely the set of integers, the set of rational numbers etc. We would like to enrich the structure of group by attaching some additional properties to it. The motivation arises from the fact that integers follow a definite pattern with respect to the ordinary addition and multiplication. In this way we are led to the concept of a ring which we define as follows:

Definition 7.1. Ring.

A system $\langle R, +, . \rangle$ where R is a non-void set, $+$ and $.$, two binary compositions defined on the set R , is called a *Ring* if it satisfies the following postulates.

For any $a, b, c \in R$

$$R-1. (a+b)+c=a+(b+c).$$

$$R-2. a+b=b+a.$$

$$R-3. \text{ There exists } 0 \in R \text{ such that } a+0=a.$$

$$R-4. \text{ For each } a \in R, \exists -a \in R \text{ such that } a+(-a)=0.$$

$$R-5. (a.b).c=a.(b.c).$$

$$R-6. a.(b+c)=a.b+a.c$$

$$(b+c).a=b.a+c.a.$$

(Left distributive law)

(Right distributive law)

REMARKS. 1. $+$ and $.$ will be called *addition* and *multiplication* respectively.

2. Postulates $R-1$ to $R-4$ show that a ring R is an Abelian additive group. Thus the additive identity 0 of R is unique. We call it, the *zero* of the ring R . Also because of the same reason, for each $a \in R$, $-a$ is unique; we call it the *additive inverse* or *negative* of a .

3. For any two elements a, b of a ring R , we shall denote $a+(-b)$ by $a-b$.

4. For convenience sake we shall usually write ab for $a.b$.

5. Following the convention set up for groups, if there is no confusion regarding the underlying binary compositions in a ring $\langle R, +, \cdot \rangle$, we shall say ' R is a ring'. Thus the phrase ' R is a ring' means that R is a non-void set with two binary compositions $+$ and $\cdot$, such that $\langle R, +, \cdot \rangle$ is a ring.

Definition 7.2. A ring R in which $ab=ba$ for every $a, b \in R$, is called a *commutative ring*.

Definition 7.3. An element e of a ring R is called a *unity* (or an *identity*) of R if $ae=ea=a$ for every $a \in R$.

In general a ring may or may not have a unity, however it can be easily shown that if a ring R has an element e such that $ae=ea=a$ for all $a \in R$ then e is unique and this e is called the *unity* or the *identity* of R . It is customary to denote the unity of a ring R by 1.

REMARK. It is clear from the definition that even if a ring R has unity e , e may not be different from 0. However if R has at least two elements, we shall see in due course that if R has unity then it must be different from zero.

EXAMPLE 1. The set $\mathbf{Z}$ is a ring under usual addition and multiplication. It is a commutative ring with unity. The integer *one* is the unity of $\mathbf{Z}$. Also each of the sets $\mathbf{Q}, \mathbf{R}, \mathbf{C}$, of all rational numbers, all real numbers and all complex numbers respectively is a commutative ring with unity under usual addition and multiplication.

EXAMPLE 2. The set E of even integers is a ring under usual addition and multiplication. E is a commutative ring without unity since there is no even integer e such that $ey=ye=y$ for all even integers y .

EXAMPLE 3. The set M of 2×2 matrices over integers form a ring under matrix addition and multiplication. M is a non-commutative ring with unity. The zero matrix is the additive identity and the matrix $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ is the unity. For $A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$, $B = \begin{pmatrix} -a_{11} & -a_{12} \\ -a_{21} & -a_{22} \end{pmatrix}$ is the negative of A .

To show that M is non-commutative it is sufficient to find two matrices A and $B \in M$ such that $AB \neq BA$. Take $A = \begin{pmatrix} 1 & 2 \\ 0 & 3 \end{pmatrix}$ and $B = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}$ then $AB = \begin{pmatrix} 3 & 0 \\ 3 & 0 \end{pmatrix}$ and $BA = \begin{pmatrix} 1 & 2 \\ 1 & 2 \end{pmatrix}$. Clearly $AB \neq BA$.

EXAMPLE 4. Consider the set $X = \{0, 1, 2, 3, 4, 5\}$. X is a ring under addition and multiplication modulo 6 (see Chapter 1)

EXAMPLE 5. Let G be an additive Abelian group with at least two elements. Define a binary composition on the elements of G as under.

For any $x, y \in G$ set $x \cdot y = 0$.

Then $\langle G, +, \cdot \rangle$ is a ring, G is a commutative ring without unity.

EXAMPLE 6. Let F denote the set of all mappings $f: \mathbf{R} \rightarrow \mathbf{R}$ such that f is continuous on $\mathbf{R}$.

For $f, g \in F$ we define their sum and product as follows:

$$(f+g)(x) = f(x) + g(x) \text{ for all } x \in \mathbf{R}.$$

$$(f \cdot g)(x) = f(x)g(x) \text{ for all } x \in \mathbf{R}.$$

Then $\langle F, +, \cdot \rangle$ is a commutative ring with unity whose zero element is a map $w: \mathbf{R} \rightarrow \mathbf{R}$ such that $w(x) = 0 \forall x \in \mathbf{R}$. The additive inverse of $f \in F$ is a map $f': \mathbf{R} \rightarrow \mathbf{R}$ such that $f'(x) = -f(x) \forall x \in \mathbf{R}$. That F is a commutative ring is seen as under.

For any $f, g \in F$, $(f \cdot g)(x) = f(x)g(x)$ and $(g \cdot f)(x) = g(x)f(x)$. But $f(x) \cdot g(x) = g(x) \cdot f(x)$ in which multiplication is commutative; therefore, $f(x)g(x) = g(x)f(x)$; hence $(f \cdot g)(x) = (g \cdot f)(x) \forall x \in \mathbf{R}$. Thus $f \cdot g = g \cdot f$ for all $f, g \in F$.

Finally let $\theta: \mathbf{R} \rightarrow \mathbf{R}$ be the mapping defined by $\theta(x) = 1 \forall x \in \mathbf{R}$. Then $(\theta \cdot f)(x) = f(x) = (f \cdot \theta)(x) \forall x \in \mathbf{R}$ gives $\theta \cdot f = f \cdot \theta = f$. Hence θ is unity of the ring F .

2. Preliminary Results

Theorem 7.4. For all a, b in a ring R

(i) $a0 = 0a = 0$.

(ii) $a(-b) = (-a)b = -ab$.

(iii) $(-a)(-b) = ab$.

Proof: (i) For all $a \in R$, $a0 = a(0+0)$

$$= a0 + a0.$$

—R-3

$$a0 + (-a0) = (a0 + a0) + (-a0)$$

$$0 = a0 + [a0 + (-a0)]$$

$$= a0 + 0$$

$$= a0.$$

Similarly

$$0a = 0.$$

Hence

$$a0 = 0a = 0 \quad \forall a \in R.$$

(ii) $a[b + (-b)] = a0 = 0$

—[By (i) above]

$$\Rightarrow ab + a(-b) = 0$$

$$\Rightarrow a(-b) = -(ab).$$

Similarly we can prove that $(-a)b = -ab$.

(iii) $(-a)(-b) = -[a(-b)] = -(-ab)$

[By part (ii) above]

$$= ab. \blacksquare$$

REMARK 1. If R is a ring with unity 1 and $1 = 0$ then $R = (0)$.

Proof: Let $x \in R$ then $x = 1x = 0x = 0$ by (i) above, so $R = (0)$. $\blacksquare$

Thus if a ring R with unity 1 has at least two elements then $1 \neq 0$.

So as to avoid the trivial situation, when we say that R is a ring

with unity then this would always mean that its unity 1 is not equal to zero.

REMARK 2. (i) By nx we mean $x+x+\dots$, n times if n is a positive integer and $(-x)+(-x)+\dots$, $(-n)$ times if n is a negative integer.

(ii) If n is a positive integer, by x^n we mean $xxx\dots$, n factors.

Thus $5x = x+x+x+x+x \quad \forall x \in R,$
 $(-5)x = (-x)+(-x)+(-x)+(-x)+(-x) \quad \forall x \in R$
 and $x^7 = xxxxxxx \quad \forall x \in R.$

3. Special Kinds of Rings

We know that given two integers a, b ; $ab=0$ implies $a=0$ or $b=0$. Thus in general if a, b are elements of a ring then one expects that whenever $ab=0$ then one should get $a=0$ or $b=0$. However this is not true. For let

$A = \begin{pmatrix} 0 & 3 \\ 0 & 9 \end{pmatrix}, B = \begin{pmatrix} 0 & 4 \\ 0 & 0 \end{pmatrix}$
 then $AB = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$ while A and B are non-zero matrices belonging to the ring M of 2×2 matrices over integers. Thus it is possible that in a ring R , there may exist two non-zero elements, a, b such that $ab=0$.

Definition 7.5. (zero-divisors).

Let R be a ring, an element $a \in R$ is called a *left zero-divisor* if $ab=0$ for some non-zero $b \in R$. An element $a \in R$ is called a *right zero-divisor* if there exists a non-zero $b \in R$ such that $ba=0$.

Trivially if a ring R contains at least two elements, then 0 is a left as well as right zero-divisor of R . 0 is called a *trivial zero-divisor*. A non-zero element of R which is a left (right) zero-divisor is called a *proper left (right) zero-divisor*. A non-zero element of R which is either left or a right zero-divisor is called a *proper zero-divisor* of R . If R contains no proper right or left zero-divisors, we say that ' R is a ring without zero-divisors'.

Definition 7.6. A commutative ring R which contains no proper zero-divisors is called an *integral domain*.

Alternatively a commutative ring R is called an *integral domain* if for all $a, b \in R$, $ab=0 \Rightarrow a=0$ or $b=0$.

EXAMPLE 7. The set $\mathbb{Z}$ under usual addition and multiplication is an integral domain, because for any two integers a, b ; $ab=0 \Rightarrow a=0$ or $b=0$.

EXAMPLE 8. Consider the ring $X = \{0, 1, 2, 3, 4, 5, 6, 7\}$ under the addition and multiplication modulo 8. This ring is commutative

however it is not an integral domain. Since $2 \in X$, $4 \in X$ are two non-zero elements such that $2 \cdot 4 \equiv 0$.

EXAMPLE 9. The set $\mathbb{Q}$ under ordinary addition and multiplication is a integral domain, for $ab=0 \Rightarrow a=0$ or $b=0 \forall a, b \in \mathbb{Q}$.

REMARK. In later chapters, we shall be using frequently the word 'domain' for integral domain.

Definition 7.7. An element a of a ring R with unity 1 is said to be unit (or invertible) if there exists an element b of R such that $ab=ba=1$.

EXAMPLE 10. In example 8 above, 5 is a unit of the ring X , since $5 \cdot 5 \equiv 1 \pmod{8}$. 1, 3 and 7 are also units as $3 \cdot 3 \equiv 1 \pmod{8}$, $1 \cdot 1 \equiv 1 \pmod{8}$, and $7 \cdot 7 \equiv 1 \pmod{8}$, while 0, 2, 4 and 6 are not units.

REMARK. An element a of a ring R with unity is called a *non-unit* if it is not a unit. Thus in example 8; 0, 2, 4 and 6 are non-units.

EXAMPLE 11. Consider ring $R=\{0, 1, 2, 3, 4, 5, 6\}$ under addition and multiplication modulo 7. Since $1 \cdot 1 \equiv 1 \pmod{7}$, $2 \cdot 4 \equiv 1 \pmod{7}$, $3 \cdot 5 \equiv 1 \pmod{7}$ and $6 \cdot 6 \equiv 1 \pmod{7}$, we see that all non-zero elements of R are invertible.

Theorem 7.8. In a ring R with unity 1, units form a subgroup of the semi-group $\langle R, \cdot \rangle$.

Proof: Let S be the set of units. Clearly $1 \cdot 1 = 1$ gives $1 \in S$. Consider $a, b \in S$. Then as a, b are units $\exists c, d \in S$ such that $ac=ca=1$ and $bd=db=1$. Then $(ab)(dc)=a(bd)c=a \cdot 1 \cdot c=ac=1$. Similarly $(dc)(ab)=1$. Thus ab is a unit; so $ab \in S$. Clearly $c=a^{-1}$ is a unit. Thus $a^{-1} \in S$. Hence S is a subgroup of $\langle R, \cdot \rangle$. ■

Definition 7.9. (Division ring).

A ring with unity in which all non-zero elements form a group under multiplication is called a *division ring*.

Definition 7.10. (Field).

A commutative division ring is called a *field*.

EXAMPLE 12. Consider M , the set of all 2×2 matrices of the type $\begin{pmatrix} a & b \\ -\bar{b} & \bar{a} \end{pmatrix}$ where a, b are complex numbers and $\bar{a}, \bar{b}$ are their conjugates. M is a ring with unity $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ under matrix addition and multiplication. Let A be a non-zero matrix in M . Then $A = \begin{pmatrix} x+yi & u+vi \\ -u+vi & x-yi \end{pmatrix}$ where x, y, u and v are not all zero.

Consider $B = \begin{pmatrix} \frac{x-yi}{x^2+y^2+u^2+v^2} & -\frac{u+vi}{x^2+y^2+u^2+v^2} \\ \frac{u-vi}{x^2+y^2+u^2+v^2} & \frac{x+yi}{x^2+y^2+u^2+v^2} \end{pmatrix}$

Clearly $B \in M$ and $AB = BA = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$.

Thus every non-zero matrix of M is a unit. Hence M is a division ring. We show that M is not a field

Now $\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix} \in M$

but $\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix} = \begin{pmatrix} 0 & -i \\ -i & 0 \end{pmatrix}$

and $\begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix} \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} = \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix}$.

So $\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix} \neq \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix} \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$.

EXAMPLE 13. In example 11 we have seen that R is a commutative ring with unity and each of its non-zero elements is invertible, therefore R is a field.

EXAMPLE 14. The set $\mathbf{C}$ of all complex numbers is a field under usual addition and multiplication. $\mathbf{C}$ is clearly commutative and 1 is unity of $\mathbf{C}$. Also for every non-zero complex number $a+bi$ its reciprocal $\frac{a-bi}{a^2+b^2}$ works as its multiplicative inverse.

Similarly since reciprocal of a non-zero rational number is a rational number and reciprocal of a non-zero real number is a real number, $\mathbf{Q}$ and $\mathbf{R}$ are also fields.

Theorem 7.11. A commutative ring R is without zero-divisors if and only if for all $a, b, c \in R$, $ab=ac$ and $a \neq 0$ imply $b=c$.

Proof: Let R be without zero-divisors and let $a, b, c \in R$ be such that $ab=ac$ and $a \neq 0$. Now $ab=ac \Rightarrow a(b-c)=0$. As $a \neq 0$ and R has no non-zero zero-divisors, $b-c=0$ i.e. $b=c$.

Conversely suppose that R satisfies the given condition and let $a, b \in R$ be such that $ab=0$ and $a \neq 0$. Now $0=a0$ gives $ab=a0$ with $a \neq 0$. This by our hypothesis implies $b=0$. Hence R is a ring without zero-divisors. ■

Definition 7.12. (Cancellation laws).

A ring R is said to satisfy the *left cancellation law* if for all $a, b, c \in R$, $a \neq 0$, $ab=ac \Rightarrow b=c$. A similar definition can be given for *right cancellation law*.

REMARK. It should be emphasized here that by cancellation in ring we always mean cancellation of non-zero elements.

Theorem 7.13. Any finite non-zero integral domain is a field.

Proof: Let $R = \{a_1, a_2, \dots, a_n\}$ be a non-zero finite integral domain with $a_1=0$ and $n \geq 2$. Consider any $a (\neq 0) \in R$, then

$$aa_1, aa_2, aa_3, \dots, aa_n$$

are all n distinct elements of R , as cancellation laws hold in R . Consequently given any $a_k \in R$, a_k is one of aa_i 's ($i=1, 2, \dots, n$) i.e.

$$a_k = aa_i \text{ for some } i. \quad \dots (1)$$

Thus in particular there exists $a_i \in R$ such that

$$a = aa_i. \quad \dots (2)$$

Then for any $b \in R$, $ab = (aa_i)b = a(a_ib)$; hence $b = a_ib$. Since R is commutative, this in turn implies that a_i is the unity of R . Let us denote a_i by e . Now (1) gives there exists $a \in R$ such that $e = aa_i$.

So $a^{-1} = a_j \in R$. Hence R is a field. ■

REMARK. Another proof of the above theorem can be given by using the fact that a finite semi-group with cancellation laws is a group. Here $R - \{0\}$ is a finite semi-group satisfying both left and right cancellation laws.

Theorem 7.13. (a) Any finite non-zero ring R without zero-divisors is a division ring.

Proof: The proof is exactly similar to the proof of Theorem 7.13 except that after equation (2) we proceed as under.

$$a = aa_i \Rightarrow a(a_i a - a) = a^2 - a^2 = 0 \Rightarrow a_i a = a \text{ as } a \neq 0.$$

Now for any $x \in R$, $ax = (aa_i)x = a(a_i x) \Rightarrow x = a_i x$ and further $xa = x(a_i a) = (xa_i)a \Rightarrow x = xa_i$. Hence a_i is the unity of R . Finally (1) gives $a_i = aa_j$ for some $a_j \in R$. Also $a(a_j a - a_i) = (aa_j)a - aa_i = a_i a - aa_i = 0 \Rightarrow a_j a = a_i$. Consequently R is a division ring. ■

REMARK. By a famous theorem of Wedderburn, "every finite division ring is a field". Hence we can say that 'any finite non-zero ring without zero-divisors is a field'.

WORKED-OUT EXERCISES

Exercise 1. Let R be a system satisfying all the postulates for a ring with the possible exception of $a+b=b+a$. If there exists one element $c \in R$ such that $ac=bc \Rightarrow a=b \forall a, b \in R$, prove that R is a ring.

Solution. $(a+b)(c+c) = a(c+c) + b(c+c) = ac + (ac+bc) + bc$

$$\text{Again } (a+b)(c+c) = (a+b)c + (a+b)c = ac + (bc+ac) + bc$$

$$\text{These give } ac + bc = bc + ac$$

$$\Rightarrow (a+b)c = (b+a)c$$

$$\Rightarrow a+b = b+a \text{ by the property of element } c.$$

Hence R is a ring.

Exercise 2. If in a ring R , $x^3=x$ for all x in R , then show that R is commutative.

Solution. For all $x, y \in R$, $(x^2y - x^2yx^2)^2 = x^2yx^2y - x^2yx^2yx^2 - x^3yx^4y + x^2yx^4yx^2 = 0$ as $x^4=x^2$.

$$\text{Thus } x^2y - x^2yx^2 = (x^2y - x^2yx^2)^3 = 0 \Rightarrow x^2y = x^2yx^2.$$

Similarly we get $yx^2 - x^2y = 0$. Hence $x^2y = yx^2$ (1)

Further $(x^2 - x)^3 = x^3 - x \Rightarrow -3x + 3x^2 = 0 \Rightarrow 3x = 3x^2$... (2)

Finally $(x^2 - x)^2 = x^4 - 2x^3 + x^2 = 2x^2 - 2x^3 = 2x^2 - 2x = x - x^2$ by (2)

But now by (1), $(x^2 - x)^2 y = y(x^2 - x)^2$ so we get $(x - x^2)y = y(x - x^2)$... (3)

Again by (1) $x^2y = yx^2$. This together with (3) imply that $xy = yx$. Hence R is commutative.

Exercise 3. If in a ring R with unity, $(xy)^2 = x^2y^2$ for all $x, y \in R$, then R is commutative.

Solution. For all $x, y \in R$, $[x(y+1)]^2 = x^2(y+1)^2$

$$\Rightarrow (xy+x)^2 = x^2(y^2+2y+1)$$

$$\Rightarrow (xy)^2 + xyx + x^2y + x^2 = x^2y^2 + 2x^2y + x^2$$

$$\Rightarrow xyx = x^2y \quad \dots (1)$$

Replacing x by $x+1$ we get that $(x+1)y(x+1) = (x+1)^2y$

$$\Rightarrow (xy+y)(x+1) = (x^2+2x+1)y$$

$$\Rightarrow xyx + xy + yx + y = x^2y + 2xy + y$$

$$\Rightarrow yx = xy; \text{ since by (1) } xyx = x^2y.$$

Hence R is a commutative ring.

Exercise 4. Let $S = \{g_1, g_2, \dots, g_n\}$ be a finite semi-group with composition denoted multiplicatively. Let R be a ring.

We define $R(S) = \{ \sum_{i=1}^n a_i g_i \mid a_i \in R, g_i \in S \}$ where $\sum_{i=1}^n a_i g_i$ is just a formal symbol.

Thus $\sum a_i g_i = \sum b_i g_i \Leftrightarrow a_i = b_i \forall i = 1, 2, \dots, n$.

We define $+$ and $\cdot$ on $R(S)$ as under

$$\sum_{i=1}^n a_i g_i + \sum_{i=1}^n b_i g_i = \sum_{i=1}^n (a_i + b_i) g_i$$

$$\left(\sum_{i=1}^n a_i g_i \right) \left(\sum_{j=1}^n b_j g_j \right) = \sum_{k=1}^n c_k g_k; \text{ where } c_k = \sum a_i b_j; \text{ the}$$

sum being taken over all pairs a_i, b_j such that $g_i g_j = g_k$. Verify that $R(S)$ is a ring.

Note. $R(S)$ is called semi-group ring.

Solution. Clearly $(\sum_i a_i g_i + \sum_i b_i g_i) \cdot \sum_i c_i g_i = \sum_i (a_i + b_i) g_i \cdot \sum_i c_i g_i$

$$= \sum_i [(a_i + b_i) + c_i] g_i$$

$$= \sum_i [a_i + (b_i + c_i)] g_i = \sum_i a_i g_i + \sum_i (b_i + c_i) g_i$$

$$= \sum_i a_i g_i + (\sum_i b_i g_i + \sum_i c_i g_i)$$

$$\text{and } \sum_i a_i g_i + \sum_i b_i g_i = \sum_i (a_i + b_i) g_i = \sum_i (b_i + a_i) g_i = \sum_i b_i g_i + \sum_i a_i g_i$$

Now $0 = \sum_i d_i g_i$ where $d_i = 0 \forall i$, works as zero for $R(S)$ as

$$\sum_i a_i g_i + \sum_i d_i g_i = \sum_i (a_i + d_i) g_i = \sum_i a_i g_i \text{ as } d_i = 0 \forall i.$$

$$\text{Also } \sum_i (-a_i) g_i + \sum_i a_i g_i = \sum_i (-a_i + a_i) g_i = 0.$$

Thus $\langle R(S), + \rangle$ is an Abelian group.

We denote, by $\sum_{g_i g_j = g_k} a_i b_j$ the element $c_k = \sum a_i b_j$ of ring R ,

where the summation is taken over all a_i, b_j such that $g_i g_j = g_k$. Similarly $\sum_{(g_i g_j) g_k = g_u} a_i b_j c_k$ and $\sum_{g_i (g_j g_k) = g_v} a_i b_j c_k$ stand for the elements

$\alpha_u = \sum a_i b_j c_k, \beta_v = \sum a_i b_j c_k$ of the ring R respectively where the two summations run over triples a_i, b_j, c_k such that $(g_i g_j) g_k = g_u$ and $g_i (g_j g_k) = g_v$, respectively.

Now $[(\sum_i a_i g_i)(\sum_j b_j g_j)](\sum_l c_l g_l) = (\sum_k \alpha_k g_k)(\sum_l c_l g_l)$ [where

$$\alpha_k = \sum_{g_i g_j = g_k} a_i b_j] = \sum_u \beta_u g_u$$

$$\begin{aligned} \text{where } \beta_u &= \sum_{g_k g_l = g_u} \alpha_k c_l = \sum_{g_k g_l = g_u} \left(\sum_{g_i g_j = g_k} a_i b_j \right) c_l \\ &= \sum_{(g_i g_j) g_l = g_u} a_i b_j c_l \end{aligned} \quad \dots (1)$$

Again $(\sum_i a_i g_i)[(\sum_j b_j g_j)(\sum_l c_l g_l)] = (\sum_i a_i g_i) \sum_k \beta_k g_k$, [where

$$\begin{aligned} \beta_k &= \sum_{g_j g_l = g_k} b_j c_l = \sum_v r_v g_v \text{ where } r_v = \sum_{g_i g_k = g_v} a_i \beta_k = \sum_{g_i g_k = g_v} a_i \left(\sum_{g_j g_l = g_k} b_j c_l \right) \\ &= \sum_{g_i (g_j g_l) = g_v} a_i b_j c_l. \end{aligned}$$

Since S is a semi-group under multiplication, $g_v = g_u \forall u, v = 1, 2, \dots, n$. Hence $[(\sum_i a_i g_i)(\sum_j b_j g_j)](\sum_l c_l g_l)$

$$= \sum_i a_i g_i [(\sum_j b_j g_j)(\sum_l c_l g_l)].$$

Thus $\langle R(S), \cdot \rangle$ is a semi-group.

Finally $\sum_j a_i g_i (\sum_j b_j g_j + \sum_j c_j g_j)$

$$= \sum_i a_i g_i (\sum_j (b_j + c_j) g_j)$$

$$= \sum_i a_i g_i (\sum_j d_j g_j) \text{ when } d_j = b_j + c_j \forall j = 1, 2, \dots, n$$

$$= \sum_k \alpha_k g_k, \text{ where } \alpha_k = \sum_{i,j} a_i d_j = \sum_{i,j} a_i (b_j + c_j)$$

$$= \sum_{g_i g_j = g_k} a_i b_j + \sum_{g_i g_j = g_k} a_i c_j$$

$$= \beta_k + \gamma_k \text{ where } \beta_k = \sum_{g_i g_j = g_k} a_i b_j, \gamma_k = \sum_{g_i g_j = g_k} a_i c_j$$

$$\begin{aligned}
 \text{Again } & (\sum_i a_i g_i)(\sum_j b_j g_j) + (\sum_i a_i g_i)(\sum_j c_j g_j) \\
 &= \sum_k \beta_k g_k + \sum_k \gamma_k g_k \\
 &= \sum_k (\beta_k + \gamma_k) g_k \\
 &= \sum_k \alpha_k g_k
 \end{aligned}$$

Hence $\sum a_i g_i (\sum b_j g_j + \sum c_j g_j) = (\sum a_i g_i)(\sum b_j g_j) + (\sum a_i g_i)(\sum c_j g_j)$

Similarly it can be proved that

$$(\sum_j b_j g_j + \sum_j c_j g_j)(\sum_i a_i g_i) = (\sum_j b_j g_j)(\sum_i a_i g_i) + (\sum_j c_j g_j)(\sum_i a_i g_i).$$

Consequently $\langle R(S), +, \cdot \rangle$ is a ring.

Exercise 5. (Kaplansky) Prove, that if an element of a ring with unity has more than one right inverse, then the set of its right inverses is not finite.

Solution. Let R be a ring with unity 1 and let $a \in R$ has more than one right inverse.

Consider the set S of right inverses of a . In other words $S = \{x \in R \mid ax = 1\}$.

Let s be a fixed element of S .

Put $T = \{xa - 1 + s \mid x \in S\}$.

$$\begin{aligned}
 \text{Since } a(xa - 1 + s) &= (ax)a - a + as \\
 &= 1a - 1 + as \\
 &= as = 1, \text{ we get } xa - 1 + s \in S, \text{ hence } T \subseteq S.
 \end{aligned}$$

Define $f: S \rightarrow T$ by $f(x) = xa - 1 + s$, $\forall x \in S$. Clearly f is onto. Further f is 1-1 since $f(x) = f(y) \Rightarrow xa - 1 + s = ya - 1 + s$

$$\Rightarrow xa = ya \Rightarrow (xa)x = (ya)x \Rightarrow x(ax) = y(ax) \Rightarrow x = y.$$

So if S is a finite set, then $|S| = |T|$. As a consequence $T = S$.

Since $s \in S$, $s \in T$ and so $s = za - 1 + s$ for some $z \in S$.

$\Rightarrow za = 1$. But $z \in S \Rightarrow az = 1$. This implies that a is unit and as such it must have exactly one right inverse. This is against the hypothesis. Hence S is not a finite set.

Exercise 6. A non-void set R with binary compositions $+$ and $\cdot$ satisfying all the postulates of a ring except right and left distributive laws, is called a Pseudo ring if $(a+b) \cdot (c+d) = a \cdot c + a \cdot d + b \cdot c + b \cdot d$ for all $a, b, c, d \in R$. Show that if $\langle R, +, \cdot \rangle$ is a pseudo-ring and $z = 0 \cdot 0$ then $3z = 0$ and $a \cdot 0 = 0 \cdot a = a \cdot z = z \cdot a = z$ for all $a \in R$. Further if we define $*$ on R by $x * y = x \cdot y - z$ for all $x, y \in R$ then prove that $\langle R, +, * \rangle$ is a ring and $x * z = z * x = 0$ for all $x \in R$.

Solution. $0 \cdot 0 = (0 + 0) \cdot (0 + 0) = 0 \cdot 0 + 0 \cdot 0 + 0 \cdot 0 + 0 \cdot 0 \Rightarrow z = 4z \Rightarrow 3z = 0$. Further $a \cdot 0 = (a + 0) \cdot (0 + 0) = a \cdot 0 + a \cdot 0 + z + z \Rightarrow a \cdot 0 + 2z = 0$. $\Rightarrow a \cdot 0 + 3z = z \Rightarrow a \cdot 0 = z$. Similarly it can be proved that $0 \cdot a = z$.

In particular this implies $a^2 \cdot 0 = 0 \cdot a^2 = z$.

Thus $a \cdot z = a \cdot (a \cdot 0) = (a \cdot a) \cdot 0 = a^2 \cdot 0 = z$. Similarly it follows that $z \cdot a = z$.

Lastly for all $a, b, c \in R$.

$$\begin{aligned}(a * b) * c &= (a \cdot b - z) * c = (a \cdot b - z) \cdot c - z \\ &= (a \cdot b - z) \cdot (c + 0) - z \\ &= (a \cdot b) \cdot c + (a \cdot b) \cdot 0 - z \cdot c - z \cdot 0 - z \\ &= (a \cdot b) \cdot c + z - z - z - z \\ &= (a \cdot b) \cdot c + z \text{ as } 3z = 0.\end{aligned}$$

Again $a * (b * c) = a * (b \cdot c - z)$

$$\begin{aligned}&= a \cdot (b \cdot c - z) - z \\ &= (a + 0) \cdot (b \cdot c - z) - z \\ &= a \cdot (b \cdot c) - a \cdot z + 0 \cdot (b \cdot c) - 0 \cdot z - z \\ &= a \cdot (b \cdot c) - z + z - z - z \\ &= a \cdot (b \cdot c) + z.\end{aligned}$$

Since $(a \cdot b) \cdot c = a \cdot (b \cdot c)$, we get that $*$ is associative.

Further $a * (b + c) = a \cdot (b + c) - z$

$$\begin{aligned}&= (a + 0) \cdot (b + c) - z \\ &= a \cdot b + a \cdot c + 0 \cdot b + 0 \cdot c - z \\ &= a \cdot b + a \cdot c + z + z - z = a \cdot b + a \cdot c + z\end{aligned}$$

Also $(a * b) + (a * c) = a \cdot b - z + a \cdot c - z$

$$\begin{aligned}&= a \cdot b + a \cdot c - 3z + z \\ &= a \cdot b + a \cdot c + z \text{ as } 3z = 0.\end{aligned}$$

Hence $a * (b + c) = a * b + a * c$

Similarly it can be proved that $(b + c) * a = b * a + c * a$.

Consequently $\langle R, +, * \rangle$ is a ring.

Finally $x * z = x \cdot z - z = z - z = 0$ and $z * x = z \cdot x - z = z - z = 0$.

Exercise 7. S is a non-empty set and R is a ring. Let X be the set of all functions from S to R . For $f, g \in X$ define $f + g$ and $f \cdot g$ as under:

For all $x \in X$, $(f + g)(x) = f(x) + g(x)$
and $(f \cdot g)(x) = f(x)g(x)$.

Show that under these binary compositions X is a ring.

Further if R is commutative, then X is commutative.

Solution. The function $w : S \rightarrow R$ given by $w(s) = 0 \forall s \in S$ is definitely in X , so X is a non empty set.

Clearly if $f, g \in X$, $f + g$ and $f \cdot g$ given by

$(f + g)(x) = f(x) + g(x)$ and $(f \cdot g)(x) = f(x)g(x) \forall x \in X$, are also in X .

Let $x \in S$ and $f, g, h \in X$.

Then $\{(f + g) + h\}(x) = (f + g)(x) + h(x)$

$$\begin{aligned}&= \{f(x) + g(x)\} + h(x) \\ &= f(x) + \{g(x) + h(x)\} \\ &= \{f + (g + h)\}(x)\end{aligned}$$

$$\Rightarrow (f+g)+h=f+(g+h)$$

Also $(f+g)(x)=f(x)+g(x)=g(x)+f(x)=(g+f)(x)$
 $\Rightarrow f+g=g+f.$

Again $(f+w)(x)=f(x)+w(x)=f(x)\Rightarrow f+w=f \quad \forall f \in X.$

Define $f_1 : S \rightarrow R$ by $f_1(s) = -f(s) \quad \forall s \in S.$

Then $\forall x \in S, (f+f_1)(x)=f(x)+f_1(x)=f(x)-f(x)=0=w(x)$
 $\Rightarrow f+f_1=w.$

Next $\forall x \in X [(f.g).h](x)=[(f.g)(x)]h(x)=[f(x)g(x)]h(x)$
 $=f(x)[g(x)h(x)]=f(x)[(g.h)(x)]$
 $=[f.(g.h)](x) \Rightarrow (f.g).h=f.(g.h)$

Finally $[f.(g+h)](x)=f(x)[(g+h)(x)]=f(x)[g(x)+h(x)]$
 $=f(x)g(x)+f(x)h(x)$
 $=(f.g)(x)+(f.h)(x)$
 $=(f.g+f.h)(x)$

$$\Rightarrow f.(g+h)=f.g+f.h$$

Similarly it can be proved that

$$(g+h).f=g.f+h.f.$$

Hence X is a ring.

Lastly if R is commutative then

$$(f.g)(x)=f(x)g(x)=g(x)f(x)=(g.f)(x) \Rightarrow f.g=g.f.$$

PROBLEMS

1. Prove that the set $\mathbf{Z}[i]=\{a+bi : a, b \in \mathbf{Z}\}$ is a ring under the usual addition and multiplication of complex numbers.

$\mathbf{Z}[i]$ is called the *ring of Gaussian integers*.

2. Prove that the set $R=\{a+b\sqrt{2}+c\sqrt{3}+d\sqrt{6} \mid a, b, c, d \in \mathbf{Q}\}$ is a field.

3. Is the set R' of all irrational numbers a ring under the ordinary operations of addition and multiplication of real numbers?

[Ans. No]

4. Prove that for any prime number p , the set $\mathbf{Q}(\sqrt{p})=\{a+b\sqrt{p} \mid a, b \in \mathbf{Q}\}$ is a ring under the operations of usual addition and multiplication of real numbers. Is this set a field?

[Ans. Yes]

5. Prove that for all a, b, c, d in a ring R

$$(a+b)(c+d)=ac+ad+bc+bd.$$

6. Prove that a ring R is commutative if and only if

$$(a+b)^2=a^2+2ab+b^2 \quad \forall a, b \in R.$$

7. Let R_n be the set of all $n \times n$ matrices whose elements are in a ring R . For any two matrices $A=(a_{ij}), B=(b_{ij})$ define $A+B=(a_{ij}+b_{ij})$ and $AB=(c_{ij})$ where $c_{ij}=\sum_{k=1}^n a_{ik}b_{kj}$. Show that under these

operations R_n is a ring. This ring is called the *ring of $n \times n$ matrices over R* .

Further show that if R has unity 1, then R_n is not commutative for any $n \geq 2$.

8. Let S be a non-void set and let F be the collection of all subsets of S . For any $A, B \in F$ define

$$A+B = A \cup B - A \cap B$$

$$\text{and } AB = A \cap B$$

Prove that F is a commutative ring with unity.

9. Let $G = \{(g_1, g_2, \dots, g_n)\}$ be a finite group with its composition denoted multiplicatively. Let R be a ring and let

$$R(G) = \{(a_1, a_2, \dots, a_n) \mid a_i \in R\}.$$

We define addition and multiplication compositions on $R(G)$ as under:

$$(a_1, a_2, \dots, a_n) + (b_1, b_2, \dots, b_n) \\ = (a_1 + b_1, a_2 + b_2, \dots, a_n + b_n)$$

and $(a_1, a_2, \dots, a_n)(b_1, b_2, \dots, b_n) = (c_1, c_2, \dots, c_n)$ where $c_j = \sum a_k b_l$, the sum being taken over all pairs a_k, b_l such that $g_k g_l = g_j$ in the group G . Verify that $R(G)$ is a ring

$R(G)$ is called the *group ring* of G over R .

10. Let $H = \{a_0 + a_1 i + a_2 j + a_3 k \mid a_0, a_1, a_2, a_3 \in \mathbf{R}\}$.

Define equality between the elements of H as under:

$$a_0 + a_1 i + a_2 j + a_3 k = b_0 + b_1 i + b_2 j + b_3 k \text{ if and only if} \\ a_0 = b_0, a_1 = b_1, a_2 = b_2 \text{ and } a_3 = b_3.$$

Define addition and multiplication compositions on H as follows:

$$(a_0 + a_1 i + a_2 j + a_3 k) = (b_0 + b_1 i + b_2 j + b_3 k) \\ = (a_0 + b_0) + (a_1 + b_1)i + (a_2 + b_2)j + (a_3 + b_3)k$$

$$\text{and } (a_0 + a_1 i + a_2 j + a_3 k)(b_0 + b_1 i + b_2 j + b_3 k) \\ = (a_0 b_0 - a_1 b_1 - a_2 b_2 - a_3 b_3) + (a_0 b_1 + a_1 b_0 + a_2 b_3 - a_3 b_2)i \\ + (a_0 b_2 + a_2 b_0 + a_3 b_1 - a_1 b_3)j + (a_0 b_3 + a_3 b_0 + a_1 b_2 - a_2 b_1)k.$$

Prove that H is a division ring.

11. If R is a system satisfying all the postulates for a ring with unity with the possible exception of $a+b=b+a$. Prove that this postulate holds and therefore R is a ring.

[Hint. Expand $(a+b)(1+1)$ in two different ways.]

12. If in a ring R , $x^2 = x \forall x \in R$, then prove that (i) $2x = 0 \forall x \in R$ and (ii) R is commutative.

13. Give an example of a non-commutative ring R without unity such that $(xy)^2 = x^2 y^2$ for all $x, y \in R$.

[Hint. Take $R = \left\{ \begin{pmatrix} a & b \\ 0 & 0 \end{pmatrix} \mid a, b \in \mathbf{Z} \right\}$]

14. Prove or disprove.

If R is a ring with unity such that $(xy)^3 = x^3y^3$ for all $x, y \in R$, then R is commutative.

15. Give an example of a pseudo-ring which is not a ring.

[Hint. Let $X = \{0, u, 2u\}$ be an additive cyclic group of order 3. For every $a, b \in X$, define $ab = u$.]

16. Let R be a ring and $e \in R$ be such that $ex = x \forall x \in R$, then e is said to be a *left unity* of R . Show that if R has a unique left unity then R has unity.

[Hint. Show that for each $x \in R$, $x + e - xe$ is also a left unity of R .]

17. Let R be a ring with unity e . If for some $x \in R$ there exists unique $y \in R$ such that $xy = e$, prove that x is invertible.

[Hint. Show that $x(e + y - yx) = e$.]

18. An element e of a ring is called an *idempotent* if $e^2 = e$. Further an element x of a ring R said to be *nilpotent* if $x^n = 0$ for some positive integer n .

Show that in a ring R ;

(i) A non-zero idempotent cannot be nilpotent.

(ii) In a ring with unity 1 and no proper zero-divisors, the only idempotents are the zero and the unity.

19. If in a commutative ring R , a and b are nilpotent elements then $a+b$ and ar for all $r \in R$ are nilpotent. Prove!

[Hint. If $a^m = 0, b^n = 0$ then $(a+b)^{m+n}$

$$= a^{m+n} + {}^{m+n}C_1 a^{m+n-1}b + \dots + \dots + b^{m+n} = 0.]$$

20. If a ring R has no non-zero nilpotent elements then for any idempotent e of R , $ex = xe \forall x \in R$. Prove!

[Hint. For any x in R , $(exe - xe)^2 = 0 = (ex - exe)^2$.]

21. Let R be a ring with more than one element. Let for each $a \in R$ there exists a unique element b in R such that $aba = a$. Prove that (i) $bab = b$ and (ii) R is a division ring.

[Hint. Show that R has no proper zero-divisors.]

22. Prove that a field is an integral domain.

23. Let p be a prime number. Let D_p be the set of rational numbers which, when expressed in lowest term, have form a/b where $b = p^k$ for some integer $k \geq 0$. Show that under ordinary addition and multiplication D_p is an integral domain.

24. Prove or disprove that there is an integral domain with six elements.

[Hint. Let D be an integral domain of order 6. Since $\langle D, + \rangle$ is a ring of order 6, the additive group of D is an Abelian group of order 6. So D has two elements a and b of order 2 and 3 respectively. Then $2ab = 0$. Now $a \neq 0, 2b \neq 0$; so $\langle D, +, \cdot \rangle$ cannot be an integral domain.]

25. Prove that any finite field must be of prime power order.

[Hint. If F is a finite field such that two distinct primes p and q divide order of F , then by Cauchy's theorem F contains two elements a and b of order p and q respectively with respect to addition, then $pab=0$.]

26. Show that the ring of integers modulo p is a field if and only if p is prime number.

4. Subrings and Ideals

In Chapter 2 we discussed the concept of subgroup H as a non-void subset of a group G closed under the binary composition on G such that H is itself a group under the induced composition. In analogy with this concept we now introduce the concept of a subring.

Definition 7.14. A non-void subset S of a ring R is called a *subring* of R if

(i) for any $a, b \in S$, $a+b \in S$ and $ab \in S$. In other words S is closed (or *stable*) under the addition and multiplication compositions in R .

(ii) S itself is a ring under the induced addition and multiplication compositions.

If S is a subring of R , then R is called an *overring* of S .

EXAMPLE 15. $\mathbf{Z}$ is a subring of $\mathbf{Q}$.

EXAMPLE 16. The set Y of all 2×2 matrices of the type $\begin{pmatrix} a & 0 \\ b & c \end{pmatrix}$ where a, b, c are integers is a subring of the ring M_2 of all 2×2 matrices over $\mathbf{Z}$.

Theorem 7.15. A non-void subset S of a ring R is a subring of S if and only if $a-b \in S$ and $ab \in S$ for all $a, b \in S$.

Proof: Let S be a subring of R and let $a, b \in S$. Then by definition of S , $a+b \in S$, $ab \in S$. Also as S is a ring under induced addition and multiplication, it follows that $a-b \in S$.

Conversely let S be such that for all $a, b \in S$, $a-b \in S$ and $ab \in S$.

Now $a-b \in S \forall a, b \in S$ implies that S is a subgroup of $\langle R, + \rangle$.

Thus S is closed under the addition composition in R and is an Abelian group under the induced addition. Further we are given that S is closed under multiplication. Consequently S is sub-semigroup of $\langle R, \cdot \rangle$. The laws of distributivity hold in S since they hold in R . So that S is a ring under the induced addition and multiplication. Hence S is a subring of R . ■

In complete analogy with discussion on Group Theory, it is now the turn of equivalence relation to be defined on the elements of ring R with respect to a subring S .

Definition 7.16. (i) Let R be a ring and S , a subring of R . Let $a, b \in R$, we say that a is right congruent to b modulo S [Symbolically $a \equiv_r b \pmod{S}$] if and only if $a-b \in S$ and $(a-b)d \in S$ for all $d \in R$.

(ii) Similarly we define a is left congruent to b modulo S [Symbolically $a \equiv_l b \pmod{S}$] and only if $a-b \in S$ and $d(a-b) \in S$ for all $d \in R$.

(iii) a is said to be congruent to b modulo S if a is both right and left congruent to b modulo S ; this is denoted by $a \equiv b \pmod{S}$.

Theorem 7.17. Right congruence modulo a subring S , is an equivalence relation on the elements of R .

Proof: Let $a, b, c \in R$.

(1) *Reflexivity:* Since $a-a=0$, $(a-a)d=0 \forall d \in R$ and $0 \in S$, we get $a \equiv_r a \pmod{S}$.

(2) *Symmetry:* $a \equiv_r b \pmod{S} \Rightarrow a-b \in S$ and $(a-b)d \in S \forall d \in R \Rightarrow b-a = -(a-b) \in S$ and $(b-a)d = -(a-b)d \in S \forall d \in R \Rightarrow b \equiv_r a \pmod{S}$.

(3) *Transitivity:* $a \equiv_r b \pmod{S}$, $b \equiv_r c \pmod{S} \Rightarrow a-b \in S$, $b-c \in S$, $(a-b)d \in S$ and $(b-c)d \in S \forall d \in R$.

As S is a subring, $a-c = (a-b) + (b-c) \in S$

and $(a-c)d = (a-b)d + (b-c)d \in S \forall d \in R$.

Hence $a \equiv_r c \pmod{S}$.

This proves the theorem. ■

REMARK. Similarly it can be proved that the left congruence modulo S and the congruence modulo S are equivalence relations on R .

Theorem 7.18. If S is a subring of R such that for any $a \in S$, $y \in R$, $ay \in S$ then for each $x \in R$ the subset $x+S = \{x+s \mid s \in S\}$ is the equivalence class for the relation of right congruence modulo S and conversely if for each $x \in R$, $x+S$ is the equivalence class for the relation of right congruence modulo S , then $ay \in S$ for all $a \in S$, $y \in R$.

Proof: Let the subring S be such that $ay \in S \forall a \in S, y \in R$ and let $L = \{t \in R \mid t \equiv_r x \pmod{S}\}$.

Now if $y \in L$ then $y \equiv_r x \pmod{S} \Rightarrow y-x \in S$

$\Rightarrow y \in x+S$. Consequently $L \subseteq x+S$

Again $z \in x+S \Rightarrow z-x \in S$. Then for any $y \in R$, $(z-x)y \in S$ by the hypothesis. This gives $z \equiv_r x \pmod{S}$ i.e., $z \in L$. This yields $x+S \subseteq L$.

Hence $x+S=L$.

Conversely if for each $x \in R$, $x+S$ is an equivalence class for the relation of right congruence modulo S , then as $0+S=S$ we get

$a \equiv_r 0 \pmod{S} \forall a \in S \Rightarrow (a-0)y \in S \forall y \in R \Rightarrow ay \in S \forall y \in R$. Hence $ay \in S \forall a \in S$ and $y \in R$. ■

Similarly if T be a subring of R such that for all $y \in R, t \in T, yt \in T$ then for every $x \in R, x+T$ is an equivalence class for the relation of left congruence modulo T and conversely. Further if V be a subring of R such that for all $v \in V, r \in R, vr \in V, rv \in V$ then for every $x \in R, x+V=V+x$ is the equivalence class of relation of congruence modulo V and conversely.

All that we have just discussed above motivates the following definitions.

Definition 7.19. A non-void subset I of a ring R is called a *left ideal* of R if (i) $a, b \in I \Rightarrow a-b \in I$ and (ii) $a \in I, r \in R \Rightarrow ra \in I$.

Definition 7.20. A non-void subset J of a ring R is called a *right ideal* of R if (i), $a, b \in J \Rightarrow a-b \in J$ and (ii) $a \in J, r \in R \Rightarrow ar \in J$.

Definition 7.21. A non-void subset K of a ring R is called an *ideal* (or *two-sided ideal*) if K is both left and right ideal of R . In other words $a, b \in K \Rightarrow a-b \in K$ and $a \in K, r \in R \Rightarrow ar \in K$ and $ra \in K$.

REMARK. In a ring R , (0) and R are trivially right as well as left ideals. Any right, left or two sided ideal A of R , which is not equal to (0) and R is accordingly called a *proper right, left or two-sided ideal* of R .

EXAMPLE 17. Set E of even integers is an ideal of the ring $\mathbb{Z}$ of integers as $2m-2n=2(m-n) \in E$ for all $2m, 2n \in E$; further if $r \in \mathbb{Z}$ then $r(2m)=2rm \in E \forall 2m \in E$.

EXAMPLE 18. Let R be a ring and let $a \in R$. The subset $X=\{r \in R \mid ar=0\}$ is a right ideal of R . To see this firstly note that $0 \in X$ so X is non-void. Further if $x, y \in X$ then $ax=0, ay=0 \Rightarrow a(x-y)=0 \Rightarrow x-y \in X$.

Also if $r \in R, a(xr)=(ax)r=0r=0$; hence $xr \in X$.

EXAMPLE 19. In the ring M of 2×2 matrices over integers, consider the set $L=\left\{\begin{pmatrix} a & 0 \\ b & 0 \end{pmatrix} \mid a, b \in \mathbb{Z}\right\}$.

Now $\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \in L$ and so L is non-empty.

Also $\begin{pmatrix} a & 0 \\ b & 0 \end{pmatrix} - \begin{pmatrix} c & 0 \\ d & 0 \end{pmatrix} = \begin{pmatrix} a-c & 0 \\ b-d & 0 \end{pmatrix} \in L$.

Again, if $\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \in M$ then $\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \begin{pmatrix} a & 0 \\ b & 0 \end{pmatrix} = \begin{pmatrix} \alpha a + \beta b & 0 \\ \gamma a + \delta b & 0 \end{pmatrix} \in L$.

This shows that L is a left ideal of M . However L is not a right

ideal of M since $\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \in L$ and $\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \in M$ but $\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \notin L$.

EXAMPLE 20. Let again M be the ring of 2×2 matrices over integers. Then $K = \left\{ \begin{pmatrix} a & b \\ 0 & 0 \end{pmatrix} \mid a, b \in \mathbf{Z} \right\}$

is a right ideal of M but not a left ideal of M . That K is a right ideal, can be checked easily. Now $\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \in K$, $\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \in M$, but

$$\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \notin K.$$

Thus K is not a left ideal.

REMARK. By definition every left, right or two-sided ideal is a subring but the converse is not true. This can be seen from the following example.

EXAMPLE 21. $\mathbf{Z}$ is a subring of $\mathbf{Q}$ but $\mathbf{Z}$ is not an ideal of $\mathbf{Q}$ since $\frac{1}{2} \cdot 3 = \frac{3}{2} \notin \mathbf{Z}$ though $\frac{1}{2} \in \mathbf{Q}$ and $3 \in \mathbf{Z}$.

REMARK. Let $B \subseteq A$ be two ideals of a ring R . It is immediate that if we regard A as a ring itself then B is also an ideal of A . We show by an example that the following *does not* hold:

'If $C \subseteq I \subseteq R$ be such that C is an ideal of I and I is an ideal of R then C is an ideal of R '.

EXAMPLE 22. Let $R = \left\{ \begin{pmatrix} a & b & c \\ d & e & f \\ 0 & 0 & g \end{pmatrix} \mid a, b, c, d, e, f, g \in \mathbf{Z} \right\}$
then R is a ring under matrix addition and multiplication.

Let $A = \left\{ \begin{pmatrix} 0 & 0 & x \\ 0 & 0 & y \\ 0 & 0 & 0 \end{pmatrix} \mid x, y \in \mathbf{Z} \right\},$

A is an ideal of R . (Verify !)

Let $B = \left\{ \begin{pmatrix} 0 & 0 & u \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \mid u \in \mathbf{Z} \right\}$, then B is also an ideal of A ,

since $\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \in B$, $\begin{pmatrix} 0 & 0 & u \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} - \begin{pmatrix} 0 & 0 & v \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 & u-v \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \in B$

and $\begin{pmatrix} 0 & 0 & x \\ 0 & 0 & y \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 0 & u \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 & u \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 0 & x \\ 0 & 0 & y \\ 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \in B.$

We claim that B is not an ideal of R .

To see this, consider $\begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \in R$ and $\begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \in B$.

$$\text{then } \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} \notin B.$$

Hence B is not an ideal of R .

Definition 7.22. A ring R is said to be *simple*, if

(i) $\exists a, b \in R$ such that $ab \neq 0$,

and (ii) R has no proper ideals.

REMARK. If R is a ring with unity $1 (\neq 0)$ then to prove that R is simple, we do not need show the condition (i) in the above definition, since by taking $a=1, b=1$ we see that $ab=1 \neq 0$.

Theorem 7.23. A division ring is a simple ring.

Proof: Let $A[\neq(0)]$ be an ideal of a division ring R . Since $A \neq (0), \exists a (\neq 0) \in A$. Again as R is a division ring, a is a unit in R , i.e. $\exists b \in R$ such that $ab=1$. Hence $1 \in A$, since $a \in A$. For any $r \in R$, we get $r=r1 \in A$. Consequently $R=A$. Hence R has no proper ideals. ■

REMARK. For any ring R with unity 1 , if any ideal I of R contains 1 then $I=R$. To see this consider $x \in R$ then $x=x1 \in I$ as $1 \in I$. So that $R \subseteq I$. However $I \subseteq R$. Hence $I=R$.

WORKED-OUT EXERCISES

Exercise 1. Let R be a ring with unity. If R has no right ideals except R and $\{0\}$, show that R is a division ring.

Solution. It is sufficient to prove that $\langle R - \{0\}, . \rangle$ is a group.

Let $x (\neq 0) \in R$. Consider $xR = \{xr \mid r \in R\}$. Clearly $x = x1 \in R$, so R is non-void.

Further $xy - xz = x(y - z) \in xR$ and for any $s \in R, (xr)s = x(rs) \in xR$. This implies xR is a right ideal of R . Since $x (\neq 0) \in xR$, $xR = R$.

As $1 \in R$, there exists $y \in R$ such that $xy=1$. Thus $\langle R - \{0\}, . \rangle$ is a semi-group with unity in which every element is right invertible. Hence $\langle R - \{0\}, . \rangle$ is a group. Consequently R is a division ring.

Exercise 2. Prove that $M_2 = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \mid a, b, c, d \in \mathbb{Q} \right\}$ is a simple ring.

Solution. That M_2 is a ring under matrix addition and multiplication can be easily checked. Since M_2 has unity viz. $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$,

M_2 will be simple if it has no ideals except M_2 and $\{0\}$. Let A be an ideal of M_2 different from $\{0\}$. There exists a non-zero matrix $X \in A$. Let $X = (a_{ij})$, $a_{ij} \in \mathbb{Q}$, $i, j = 1, 2$. Since $X \neq \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$, there exists $a_{rs} \in \mathbb{Q}$ such that $a_{rs} \neq 0$. Let E_{ij} denote a 2×2 matrix having 1 at i th row and j th column position and zeros elsewhere. (E_{ij} is called a matrix unit).

Put $Y = E_{1r}$ and $Z = E_{s1}$, $C = E_{2r}$, $D = E_{s2}$

$$\text{Then } YXZ + CXD = \begin{pmatrix} a_{rs} & 0 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0 & a_{rs} \end{pmatrix} = \begin{pmatrix} a_{rs} & 0 \\ 0 & a_{rs} \end{pmatrix} \in A.$$

Since $a_{rs} \neq 0$, $a_{rs}^{-1} \in \mathbb{Q}$. So the matrix $\begin{pmatrix} a_{rs}^{-1} & 0 \\ 0 & a_{rs}^{-1} \end{pmatrix} \in M_2$.

As A is an ideal of M_2 , $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} a_{rs}^{-1} & 0 \\ 0 & a_{rs}^{-1} \end{pmatrix} \begin{pmatrix} a_{rs} & 0 \\ 0 & a_{rs} \end{pmatrix} \in A$.

Hence $A = M_2$. Thus M_2 is simple.

Exercise 3. Let A be a non-empty set, and $P(A)$ denotes the ring of power set of X (compare problem 8 at the end of section 3, chapter 7). Prove that a non-empty subset I of $P(A)$ is an ideal of $P(A)$ if and only if for all $X, Y \in I$, $P(X \cup Y) \subseteq I$.

Solution. Suppose for all $X, Y \in I$, $P(X \cup Y) \subseteq I$.

$X, Y \in I \Rightarrow X \Delta Y = (X \cup Y) - (X \cap Y)$ is a subset of $X \cup Y$. This implies $X \Delta Y \in P(X \cup Y) \Rightarrow X \Delta Y \in I$.

Further for any $Z \in P(A)$, $X \cap Z \subseteq X \subseteq (X \cup Y) \Rightarrow X \cap Z \in P(X \cup Y) \Rightarrow X \cap Z \in I$. Hence I is an ideal of $P(A)$.

Conversely, let I be an ideal of $P(A)$.

Let $T \in P(X \cup Y)$. This implies that $T \subseteq X \cup Y$.

$X, Y \in I$ and I is an ideal of $P(A) \Rightarrow X \Delta Y \in I$. $X \cap Y \in I$

Further, since $T \in P(A)$, $T \cap (X \Delta Y) \in I$ and $T \cap (X \cap Y) \in I$.

But $T = [T \cap (X \Delta Y)] \Delta [T \cap (X \cap Y)]$, (Verify !)

Hence $T \in I$. As a consequence $P(X \cup Y) \subseteq I$.

PROBLEMS

1. Let R be a ring. Show that $C = \{x \in R \mid xy = yx \forall y \in R\}$ is a subring of R .

C is called the centre of R .

2. If R is a ring and $a \in R$, show that $N(a) = \{r \in R \mid ar = ra\}$ is a subring of R and that the centre of R is a subring of $N(a)$.

$N(a)$ is called the Normalizer of a in R .

3. Let A be an ideal of R such that $A \neq R$. Show that if R has unity then $1 \notin A$.

4. If A and B are ideals of a ring R , define $A : B = \{r \in R \mid rB \subseteq A\}$.

Show that $A : B$ is an ideal of R .

[Note that $rB = \{rb \mid b \in B\}$].

5. For any element a of R show that the set $Y = \{r \in R \mid ra = 0\}$ is a left ideal of R .

6. Let R be a commutative ring and let A be an ideal of R . Prove that $\sqrt{A} = \{x \in R \mid x^n \in A \text{ for some positive integer } n\}$ is an ideal of R such that (i) $A \subseteq \sqrt{A}$ (ii) $\sqrt{\sqrt{A}} = \sqrt{A}$ (iii) If R has unity and $\sqrt{A} = R$ then $A = R$.

[Hint. $a, b \in \sqrt{A} \Rightarrow a^m \in A, b^n \in A$, for integers $m, n > 0 \Rightarrow (a-b)^{m+n} = a^{m+n} - \binom{m+n}{1}a^{m+n-1}b + \dots + (-1)^{m+n}b^{m+n} \in A \Rightarrow a-b \in \sqrt{A}$]

7. Let R be a commutative ring. Prove that the set of all nilpotent elements of R is an ideal of R .

8. For any left ideal I and any right ideal J of a ring R , prove

$$I_1 = \{x \in R \mid x \equiv 0 \pmod{I}\}$$

and $J_1 = \{x \in R \mid x \equiv 0 \pmod{J}\}$ are ideals of R .

9. Show that any matrix ring F_n over a field F is a simple ring. Deduce that a simple ring may have proper one-sided ideals.

10. If I is a right ideal of a ring R , show that I_3 , the set of all 3×3 matrices with elements in I is a right ideal of R_3 , the ring of 3×3 matrices over R .

11. Prove that if R is a ring with more than one element such that $aR = R$ for every non-zero element a of R then R is a division ring.

[Hint. First show that R has no proper zero-divisors. Let c and d be non-zero elements of R . Then $cR = R, dR = R$. Thus $(cd)R = c(dR) = cR = R$. This gives $cd \neq 0$. So that R is a ring without zero-divisors. Now $c(\neq 0) \in R = cR \Rightarrow c = cu$ for some $u \in R$. Then $cu = cu^2 \Rightarrow u = u^2$ and $u \neq 0$. As $uR = R$, for any $x \in R, x = uy, y \in R \Rightarrow ux = u^2y = uy = x$. So u is a left unity of R . Now $u \in R = cR \Rightarrow \exists d \in R$ such that $u = cd$ also $u \in R = dR \Rightarrow \exists e \in R$ such that $u = de \Rightarrow e = ue = (cd)e = c(de) = cu = c$.]

12. Let E be the ring of even integers. Prove that the set all matrices of the form $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ where each a is of the type $4k$ (k is some integer) is an ideal of E_2 , the ring of 2×2 matrices over E .

13. Show that a commutative ring R with unity is a field if and only if the only ideals of R are $\{0\}$ and R itself.

14. If R is a ring such that for all $x \in R, x^2 - x \in$ the centre of R , show that R is commutative.

5. Algebra of Ideals

Theorem 7.24. *Intersection of two right (left) ideals of a ring R is a right (left) ideal of R .*

Proof: Let A and B be two right ideals of a ring R . Since $0 \in A$, $0 \in B \Rightarrow 0 \in A \cap B$, $A \cap B$ is non-void. Let $x, y \in A \cap B$ and $r \in R$ then $x, y \in A \cap B \Rightarrow x \in A$ and $x \in B$, $y \in A$ and $y \in B$. As A is a right ideal of R , $x-y \in A$ and $xr \in A \forall r \in R$. For the same reasons $x-y \in B$ and $xr \in B \forall r \in R$. Thus $x, y \in A \cap B \Rightarrow x-y \in A \cap B$ and $xr \in A \cap B \forall r \in R$.

Consequently $A \cap B$ is a right ideal of R . Similarly it can be shown that if A and B are left ideals of R then $A \cap B$ is also a left ideal of R . ■

Theorem 7.25. *Intersection of any non-void family of right (left) ideals of a ring R is a right (left) ideal of R .*

Proof: Let $F = \{A_i\}_{i \in \Lambda}$ be any non-void family of right ideals of a ring R . Let $A = \bigcap_{i \in \Lambda} A_i$. Since $0 \in A_i \forall i \in \Lambda$ we have $0 \in A$.

Consequently $A \neq \phi$. On the same lines as in the above theorem we can prove that $x \in A$, $y \in A$, $r \in R \Rightarrow x-y \in A$, $xr \in A$. Hence A is a right ideal of R . Proof for left ideals is similar. ■

REMARK. If A and B are two ideals of a ring R then $A \cup B$ may not be an ideal of R . This can be well seen in the following example:

EXAMPLE 23. $A = \{2n \mid n \in \mathbb{Z}\}$ and $B = \{3n \mid n \in \mathbb{Z}\}$ are two ideals of $\mathbb{Z}$. Consider $A \cup B$. $2 \in A$, $3 \in B \Rightarrow 2, 3 \in A \cup B$, but $3-2=1 \notin A \cup B$ as 1 is neither a multiple of 2 nor a multiple of 3. Hence $A \cup B$ is not an ideal of $\mathbb{Z}$.

One should notice that in this example "neither A is contained in B nor B is contained in A ." Now we prove:

Theorem 7.26. *For two ideals A and B of a ring R , $A \cup B$ is an ideal of R if and only if either $A \subseteq B$ or $B \subseteq A$.*

Proof: If $A \subseteq B$ then $A \cup B = B$ is an ideal of R . Again $B \subseteq A \Rightarrow A \cup B = A$ is an ideal of R .

Conversely let $A \cup B$ be an ideal of R . Suppose that on the contrary "neither A is contained in B nor B is contained in A ." Then there exist elements $x \in A-B$ and $y \in B-A$. $x \in A$, $y \in B \Rightarrow x, y \in A \cup B$. As $A \cup B$ is an ideal, this in turn implies $x-y \in A \cup B$. So either $x-y \in A$ or $x-y \in B$. In case $x-y \in A$ we get $y = x - (x-y) \in A$. This contradicts the choice of y . Similarly if $x-y \in B$ then $x = (x-y) + y \in B$ which is against the fact that $x \notin B$. Hence our supposition is wrong. In other words either $A \subseteq B$ or $B \subseteq A$. ■

We show that union of more than two ideals may be ideal even if none of them contains the union of others. Consider the following example.

EXAMPLE 24. Let $X = \{0, a, b, c\}$ be a set, subject to the conditions $2a = 2b = 2c = 0$, $a + b = b + a = c$, $a + c = c + a = b$, $b + c = c + b = a$ and $\forall x, y \in X, xy = 0$. Then X is a ring. The subsets $I = \{0, a\}$, $J = \{0, b\}$ and $K = \{0, c\}$ are three ideals of X and none of them contains the union of other two. But $X = I \cup J \cup K$, is an ideal.

Definition 7.27. Let A and B be two ideals of a ring R , then, the set $A + B = \{a + b \mid a \in A, b \in B\}$ is called the *sum* of ideals A and B .

Theorem 7.28. For any two ideals A and B of a ring R , $A + B$ is an ideal of R containing both A and B .

Proof: Clearly $0 = 0 + 0 \in A + B$, so $A + B$ is non void. Consider $x = a_1 + b_1, y = a_2 + b_2, a_1, a_2 \in A$ and $b_1, b_2 \in B$. Then $x - y = (a_1 - a_2) + (b_1 - b_2) \in A + B$ as $a_1 - a_2 \in A$ and $b_1 - b_2 \in B$. Further for any $r \in R$ $rx = a_1r + b_1r, rx = ra_1 + rb_1 \in A + B$ since $a_1r, ra_1 \in A$ and $b_1r, rb_1 \in B$. Hence $A + B$ is an ideal of R . Finally for each $a \in A$, as $a = a + 0$ and $0 \in B$ we get $a \in A + B$. Consequently $A \subseteq A + B$. Similarly $B \subseteq A + B$. This proves the theorem. ■

Definition 7.29. Let S be any subset of a ring R . Then an ideal A of R is said to be *generated* by S if (i) $S \subseteq A$ and (ii) for any ideal B of R , $S \subseteq B \Rightarrow A \subseteq B$.

We shall denote such an ideal A by the symbol $\langle S \rangle$. One can see from the definition that $\langle S \rangle$ is the intersection of all those ideals of R which contain S . Consequently $\langle S \rangle$ is unique. Further if an ideal A is generated by a singleton, say $S = \{a\}$, then A is said to be a *principal ideal* and instead of writing $A = \langle S \rangle$ we write $A = \langle a \rangle$ or (a) . Finally if for an ideal A , there exists a finite set $S \subseteq A$, which generates A , then A is said to be *finitely generated*.

Theorem 7.30. For any two ideals A and B of a ring R , $A + B = \langle A \cup B \rangle$.

Proof: By Theorem 7.28, $A + B$ is an ideal of R such that $A \subseteq A + B$ and $B \subseteq A + B$ so $A \cup B \subseteq A + B$. Let X be any ideal of R such that $A \cup B \subseteq X$. If $z \in A + B$ then $z = a + b, a \in A, b \in B$. Now $a \in A \cup B, b \in A \cup B \Rightarrow a, b \in X \Rightarrow a + b \in X \Rightarrow z \in X$.

Hence $A + B \subseteq X$. Consequently, by definition $A + B = \langle A \cup B \rangle$. ■

Definition 7.31. For any two ideals A and B of a ring R , define $AB = \{ \sum_{\text{finite}} a_i b_i \mid a_i \in A, b_i \in B \}$. AB is called the *product* of A and B .

Theorem 7.32. For any two ideals A and B of a ring R , the product AB is an ideal of R .

Proof: Clearly $0 = 00 \in AB$, so AB is non-empty. $x, y \in AB \Rightarrow$

$$x = \sum_{i=1}^n a_i b_i, y = \sum_{j=1}^m c_j d_j \text{ where } a_i, c_j \in A, b_i, d_j \in B; \text{ then } x - y = \sum_{i=1}^n a_i b_i -$$

$$\sum_{j=1}^m c_j d_j = \sum_{k=1}^{m+n} x_k y_k \text{ where } x_k = a_k \forall k=1, 2, \dots, n, x_{n+t} = -c_t \forall t=1,$$

$$2, \dots, m; y_k = b_k \forall k=1, 2, \dots, n, \text{ and } y_{n+t} = d_t \forall t=1, 2, \dots, m.$$

Since $x_k \in A, y_k \in B \forall k=1, 2, \dots, m+n$ we get $x - y \in AB$.

$$\text{Further if } r \in R \text{ then } xr = \left(\sum_{i=1}^n a_i b_i \right) r$$

$$= \sum_{i=1}^n (a_i b_i) r$$

$$= \sum_{i=1}^n a_i (b_i r) \in AB \text{ as } a_i \in A \text{ and } b_i r \in B$$

$$\forall i=1, 2, \dots, n.$$

$$\text{Similarly } rx = r \left(\sum_{i=1}^n a_i b_i \right) = \sum_{i=1}^n (r a_i) b_i \in AB, \text{ as } r a_i \in A \text{ and } b_i \in B$$

$\forall i=1, 2, \dots, n$. Hence AB is an ideal of R . ■

Definition 7.33. Powers of an ideal.

Let A be an ideal of a ring then we define $A^2 = AA$.

Thus $A^2 = \left\{ \sum_{\text{finite}} a_i b_i \mid a_i, b_i \in A \right\}$.

Using induction we can define A^n for any positive integer n .

Definition 7.34. Nilpotent ideal.

An ideal A of a ring R is said to be *nilpotent* if for some positive integer n , $A^n = (0)$.

EXAMPLE 25. Consider M_2 , the ring of all 2×2 upper triangular matrices over integers. Consider the ideal I of M_2 generated by $\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$. Clearly I is nilpotent ideal since $I^2 = (0)$.

Definition 7.35. Nil ideal.

An ideal I of a ring R is said to be a *nil ideal* if each element of I is nilpotent (i.e., for each $a \in I$, there exists a positive integer n such that $a^n = (0)$).

EXAMPLE 26. The ideal I referred to in Example 25 is clearly nil ideal since $\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}^2 = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$.

REMARK. Clearly every nilpotent ideal is nil since if A is nilpotent ideal then there exists a positive integer n such that $A^n = (0)$. So for each $a \in A$, $a^n \in A^n = (0) \Rightarrow a^n = 0$, hence A is a nil ideal. Converse is

not true, for this purpose we give following example.

EXAMPLE 27. Let Z_{p^n} ($n > 1$) be the ring of integers modulo p^n where p is a fixed prime. Let $F = \{\{\bar{a}_n\} \mid \bar{a}_n \in Z_{p^n}\}$ i.e. F is the family of all sequences $\{\bar{a}_n\}$ with $\bar{a}_n \in Z_{p^n}$.

Define addition and multiplication on F as under.

$$\{\bar{a}_n\} + \{\bar{b}_n\} = \{\bar{a}_n + \bar{b}_n\} \text{ and } \{\bar{a}_n\} \cdot \{\bar{b}_n\} = \{\bar{a}_n \bar{b}_n\}.$$

Then it can be checked that F is a ring with the sequence $\{\bar{0}_n\}$ working as zero, where $\bar{0}_n$ is zero of Z_{p^n} and $-\{\bar{a}_n\} = \{-\bar{a}_n\}$. Now let R be the set of all those sequence of F which have only *finitely many* non-zero terms. Then R is an ideal of F (Verify!).

We consider R as a ring. Let I be the ideal of R consisting of all elements of type $(p\bar{r}_1, p\bar{r}_2, \dots, p\bar{r}_n, \bar{0}, \dots)$, $\bar{r}_k \in Z_{p^k}$.

Let $a = (p\bar{r}_1, p\bar{r}_2, \dots, p\bar{r}_n, \bar{0}, \bar{0}, \dots) \in I$. Since in Z_{p^i} , $(p\bar{r}_i)^i = \bar{0}$, we get $(p\bar{r}_i)^k = \bar{0} \forall i \leq k$. Thus $a^k = (\bar{0}, \bar{0}, \bar{0}, \dots)$. Hence I is a nil ideal. We claim that I is not nilpotent. For this purpose we show that for any positive integer n there exists an element $a \in I$ such that $a^n \neq \bar{0}$.

Define $a = \{\bar{a}_k\}$ by taking $\bar{a}_{n+1} = \bar{p}$ in $Z_{p^{n+1}}$ and $\bar{a}_k = \bar{0} \forall k \neq n+1$. Thus $a = (\bar{0}, \bar{0}, \bar{0}, \dots, \bar{p}, \bar{0}, \bar{0}, \dots)$.

(n+1)th place

Since in $Z_{p^{n+1}}$, $\bar{p}^n \neq \bar{0}$, we get

$$a^n = (\bar{0}, \bar{0}, \bar{0}, \dots, \bar{p}^n, \bar{0}, \bar{0}, \dots) \neq \bar{0}$$

Hence $I^n \neq (0)$, $\forall$ integers $n > 0$. Consequently I is not a nilpotent ideal.

WORKED-OUT EXERCISES

Exercise 1. S and T are subrings of a ring R . Let $ST = \{\sum_{\text{finite}} s_i t_i, s_i \in S, t_i \in T\}$. Show that if $ST = TS$ then ST is a subring of

R . Give an example of subrings S and T of a ring R such that ST need not be a subring of R .

Solution. Clearly ST is non-empty and $x, y \in ST \Rightarrow x - y \in ST$.

Let $x, y \in ST$ then $x = \sum_{i=1}^m s_i t_i, y = \sum_{j=1}^n a_j b_j, s_i, a_j \in S$ and $t_i, b_j \in T$.

This implies that $xy = \sum_{j=1}^n \sum_{i=1}^m s_i (t_i a_j) b_j$.

Now $t_i a_j \in TS \Rightarrow t_i a_j \in ST$

$$\Rightarrow t_i a_j = \sum_{k=1}^{q(i,j)} \alpha_k^{(i)} \beta_k^{(j)}$$

where $\alpha_k^{(i)} \in S$, $\beta_k^{(j)} \in T$ and $q(i, j)$ is a natural number depending upon i and j .

$$\text{Hence } xy = \sum_{j=1}^n \sum_{i=1}^m \sum_{k=1}^{q(i,j)} s_i \alpha_k^{(i)} \beta_k^{(j)} b_j.$$

But $s_i \alpha_k^{(i)} \in S$, $\beta_k^{(j)} b_j \in T \forall k, j$ and i .

Hence $xy \in ST$.

This shows that ST is a subring of R . For second part take $R = \mathbf{Z}_3$, the ring of 3×3 matrices over $\mathbf{Z}$.

$$\text{Let } S = \left\{ \begin{pmatrix} 0 & 0 & 0 \\ a & 0 & 0 \\ b & c & 0 \end{pmatrix} \mid a, b, c \in \mathbf{Z} \right\}, T = \left\{ \begin{pmatrix} 0 & 0 & a \\ 0 & b & 0 \\ 0 & 0 & 0 \end{pmatrix} \mid a, b \in \mathbf{Z} \right\}$$

It can be verified that S and T are subrings of R .

$$\text{For any } s = \begin{pmatrix} 0 & 0 & 0 \\ x & 0 & 0 \\ y & z & 0 \end{pmatrix} \in S, t = \begin{pmatrix} 0 & 0 & \alpha \\ 0 & \beta & 0 \\ 0 & 0 & 0 \end{pmatrix} \in T$$

$$st = \begin{pmatrix} 0 & 0 & 0 \\ x & 0 & 0 \\ y & z & 0 \end{pmatrix} \begin{pmatrix} 0 & 0 & \alpha \\ 0 & \beta & 0 \\ 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & x\alpha \\ 0 & z\beta & y\alpha \end{pmatrix} \quad (1)$$

and ST is a finite sum of elements of type (1).

$$\text{Now } \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \end{pmatrix} \in S, \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \in T \Rightarrow \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix} \in ST$$

$$\text{also } \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 2 & 0 \end{pmatrix} \in S, \begin{pmatrix} 0 & 0 & 1 \\ 0 & 2 & 0 \\ 0 & 0 & 0 \end{pmatrix} \in T \Rightarrow \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 4 & 1 \end{pmatrix} \in ST$$

$$\text{So if } ST \text{ is a subring of } R, \text{ then } \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 4 & 1 \end{pmatrix} \in ST.$$

$$\text{In other words } \begin{pmatrix} 0 & 0 & 0 \\ 0 & 4 & 1 \\ 0 & 4 & 2 \end{pmatrix} \in ST.$$

This is absurd as every element of ST must be of type (1).

Hence ST is not a subring of R .

Exercise 2. Two ideals A and B of a ring R are called co-maximal $A+B=R$.

Prove that if A and B are co-maximal ideals of a commutative ring R with unity then $AB=A \cap B$.

Solution. As A is an ideal of R , $AB \subseteq AR \subseteq A$. Similarly as B is an ideal of R , $AB \subseteq B$.

Hence $AB \subseteq A \cap B$.

Let $x \in A \cap B$. Now $R = A + B \Rightarrow 1 = r + s$ for some $r \in A, s \in B$. This gives that $x = x \cdot 1 = x(r + s) = xr + xs$. But $r \in A, x \in B \Rightarrow xr = rx \in AB$ and $x \in A, s \in B \Rightarrow xs \in AB$.

As a consequence $x \in AB$. In other words $A \cap B \subseteq AB$. Hence $A \cap B = AB$.

PROBLEMS

1. Show that if A is a right ideal and B is a left ideal of a ring R , then $AB \subseteq A \cap B$.

Give an example of a ring R and ideals A and B of R such that $AB \neq A \cap B$.

2. Prove the following:

(i) If A and B are right (left) ideals of a ring R then AB is a right (left) ideal of R .

(ii) If A is a left and B is a right ideal of R then AB is a two-sided ideal of R but BA need not be even a one-sided ideal of R ; however BA is a subring of R .

(iii) If A is a left and B is a right-ideal of R then $A + B$ need not be even a one-sided ideal of R .

(iv) If A is a left ideal of R and B is a right ideal of R , $A \cap B$ need not be even a one-sided ideal of R .

[Hint. (iii) and (iv). In the ring $\mathbb{Z}_2$ of 2×2 matrices over $\mathbb{Z}$ take $B = \left\{ \begin{pmatrix} a & b \\ 0 & 0 \end{pmatrix} \mid a, b \in \mathbb{Z} \right\}$ and $A = \left\{ \begin{pmatrix} x & 0 \\ y & 0 \end{pmatrix} \mid x, y \in \mathbb{Z} \right\}$.]

3. Let A and B be ideals of a ring R such that $A \cap B = (0)$. The sum $A + B$ is called *direct sum* of A and B , and it is denoted by $A \oplus B$. Show that each element of $A \oplus B$ is *uniquely* expressible as $a + b, a \in A, b \in B$.

4. If A, B and C are ideals of a ring R prove that $A(B + C) = AB + AC$.

5. Prove that if a ring R contains a non-zero idempotent e (i.e. $e^2 = e$) such that e is not a left unity of R then R is direct sum (see problem 3 above) of non-zero right ideals eR and $S = \{a - ea \mid a \in R\}$.

6. If A, B, C are ideals of a ring such that $B \subseteq A$ prove that $A \cap (B + C) = B + (A \cap C)$.

This is called *modular law*.

Give an example to show that in general $A \cap (B + C)$ is not equal to $A \cap B + A \cap C$.

[Hint. See Example 24. Take $A = I, B = J$ and $C = K$.]

7. With reference to problem 4 of Section 4, if A, B, C are ideals of a ring R , prove the following:

$$(i) (A \cap B) : C = (A : C) \cap (B : C).$$

$$(ii) A : (B + C) = (A : B) \cap (A : C).$$

$$(iii) A : (BC) = (A : C) : B.$$

8. With reference to problem 4 of Section 4, prove the following for all ideals I, J and K of a ring R .

$$(i) I \subseteq J \Rightarrow I : K \subseteq J : K \text{ and } K : I \supseteq K : J.$$

$$(ii) I : J^{n+1} = (I : J^n) : J = (I : J) : J^n \text{ for any positive integer } n.$$

$$(iii) I : J = I : (I + J).$$

9. Let R be a ring with unity and let R_n be the ring of all $n \times n$ matrices over R . Prove the following:

(i) If I is an ideal of R then I_n , the subset of R_n consisting of those $n \times n$ matrices whose all elements are in I is an ideal of R_n .

(ii) Every ideal of R_n is of the form I_n for some ideal I of R .

(iii) If R is simple then R_n is simple.

10. Let R be a ring with the property that every subring of R is necessarily an ideal of R . If R contains no proper zero-divisors, prove that R is commutative.

11. Prove the following:

(i) The sum of two nilpotent ideals is nilpotent however the sum of two nilpotent elements may not be nilpotent.

(ii) Show that in a commutative ring R there exists a nil ideal A which contains all the nil ideals of R .

12. With reference to problem 4 of section 4, prove that if R is a ring with unity and I, J are ideals of R then $I : J = R \Leftrightarrow J \subseteq I$.

Give an example to show if R has no unity then the result need not be true in general.

[Hint. Take $R=(2), I=(4), J=(6)$].

8

HOMOMORPHISMS AND EMBEDDING OF RINGS

1. Quotient Rings, Fields and Homomorphisms

By definition, a ring R is an additive Abelian group and an ideal N of R is necessarily additive subgroup of $\langle R, + \rangle$. We can talk of the additive quotient group $R/N = \{r+N \mid r \in R\}$ of $\langle R, + \rangle$. Further notice that each $r+N$ is also an equivalence class for the relation of congruence modulo N (Theorem 7.18). Thus R/N is the set of all equivalence classes for the relation of congruence modulo N . We would like to make R/N a ring. For this purpose, we define multiplication composition as follows:

For $r+N, s+N \in R/N$, $(r+N)(s+N) = rs+N$. We notice that since R/N is an Abelian group under addition composition $\oplus$ given by $(r+N) \oplus (s+N) = (r+s)+N$, $\oplus$ is already well-defined. We verify that multiplication is also well-defined: for this purpose let

$$r+N = r_1+N \text{ and } s+N = s_1+N \text{ for some } r, r_1, s, s_1 \in R.$$

This implies $r-r_1 = n, s-s_1 = n_1$ for some $n, n_1 \in N$

$$\Rightarrow r = n + r_1 \text{ and } s = n_1 + s_1$$

$$\Rightarrow rs = nn_1 + ns_1 + r_1n_1 + r_1s_1.$$

As $n, n_1 \in N$ and N is an ideal of R , $nn_1 + ns_1 + r_1n_1 \in N$.

Consequently $rs - r_1s_1 \in N$.

Hence $rs+N = r_1s_1+N$.

i.e. $(r+N)(s+N) = (r_1+N)(s_1+N)$.

This shows that multiplication in R/N is well-defined.

Theorem 8.1. *Let R be a ring and N , an ideal of R ; then R/N is a ring under the addition and multiplication defined as under:*

For $r+N, s+N \in R/N$, $(r+N)+(s+N) = (r+s)+N$ and $(r+N)(s+N) = rs+N$.

Proof: By definition R/N is closed with respect to addition and multiplication. Let $r, s, t \in R$.

(1) $\langle R/N, + \rangle$ is an Abelian group.

(2) *Multiplication is Associative:*

$$\begin{aligned}
 (r+N)[(s+N)(t+N)] &= (r+N)(st+N) \\
 &= r(st) + N \\
 &= (rs)t + N, \text{ since } (rs)t = r(st) \text{ in } R. \\
 &= (rs+N)(t+N) \\
 &= [(r+N)(s+N)](t+N).
 \end{aligned}$$

(3) *Distributive laws:*

$$\begin{aligned}
 (r+N)[(s+N)+(t+N)] &= (r+N)[(s+t)+N] \\
 &= r(s+t) + N \\
 &= (rs+rt) + N \\
 &= (rs+N) + (rt+N) \\
 &= (r+N)(s+N) + (r+N)(t+N).
 \end{aligned}$$

Similarly we can prove that

$$[(s+N)+(t+N)](r+N) = (s+N)(r+N) + (t+N)(r+N).$$

Hence $\langle R/N, +, \cdot \rangle$ is a ring. ■

Definition. 8 2. Let R be a ring and N be any ideal of R , then the system $\langle R/N, +, \cdot \rangle$ where $R/N = \{r+N \mid r \in R\}$ and $+$ & $\cdot$ are binary compositions on R/N defined by $(r+N) + (s+N) = (r+s) + N$ and $(r+N) \cdot (s+N) = rs + N$ for all $r, s \in R$, is a ring. This ring is called the *quotient ring* of R with respect to the ideal N .

NOTATION. We shall usually denote the element $x+N$ of R/N by $\bar{x}$; thus $\bar{0}$ in this notation will mean N , the zero of R/N . One finds from definition that for any $\bar{x}, \bar{y} \in R/N$, $\overline{x+y} = \bar{x} + \bar{y}$ and $\overline{xy} = \bar{x}\bar{y}$.

EXAMPLE 1. Let $N = \{6n \mid n \in \mathbf{Z}\}$. N is an ideal of $\mathbf{Z}$. The elements of $\mathbf{Z}/N$ are the cosets $N, 1+N, 2+N, 3+N, 4+N$ and $5+N$ which according to our notation can be denoted by $\bar{0}, \bar{1}, \bar{2}, \bar{3}, \bar{4}$, and $\bar{5}$ respectively. Then $\langle \mathbf{Z}/N, +, \cdot \rangle$ is a ring where both addition and multiplication are modulo 6. See the following tables.

Addition Table

	$\bar{0}$	$\bar{1}$	$\bar{2}$	$\bar{3}$	$\bar{4}$	$\bar{5}$
$\bar{0}$	$\bar{0}$	$\bar{1}$	$\bar{2}$	$\bar{3}$	$\bar{4}$	$\bar{5}$
$\bar{1}$	$\bar{1}$	$\bar{2}$	$\bar{3}$	$\bar{4}$	$\bar{5}$	$\bar{0}$
$\bar{2}$	$\bar{2}$	$\bar{3}$	$\bar{4}$	$\bar{5}$	$\bar{0}$	$\bar{1}$
$\bar{3}$	$\bar{3}$	$\bar{4}$	$\bar{5}$	$\bar{0}$	$\bar{1}$	$\bar{2}$
$\bar{4}$	$\bar{4}$	$\bar{5}$	$\bar{0}$	$\bar{1}$	$\bar{2}$	$\bar{3}$
$\bar{5}$	$\bar{5}$	$\bar{0}$	$\bar{1}$	$\bar{2}$	$\bar{3}$	$\bar{4}$

Multiplication Table

	$\bar{0}$	$\bar{1}$	$\bar{2}$	$\bar{3}$	$\bar{4}$	$\bar{5}$
$\bar{0}$	$\bar{0}$	$\bar{0}$	$\bar{0}$	$\bar{0}$	$\bar{0}$	$\bar{0}$
$\bar{1}$	$\bar{0}$	$\bar{1}$	$\bar{2}$	$\bar{3}$	$\bar{4}$	$\bar{5}$
$\bar{2}$	$\bar{0}$	$\bar{2}$	$\bar{4}$	$\bar{0}$	$\bar{2}$	$\bar{4}$
$\bar{3}$	$\bar{0}$	$\bar{3}$	$\bar{0}$	$\bar{3}$	$\bar{0}$	$\bar{3}$
$\bar{4}$	$\bar{0}$	$\bar{4}$	$\bar{2}$	$\bar{0}$	$\bar{4}$	$\bar{2}$
$\bar{5}$	$\bar{0}$	$\bar{5}$	$\bar{4}$	$\bar{3}$	$\bar{2}$	$\bar{1}$

Definition 8.3. (Prime ideal) Let R be a commutative ring. An ideal P of R is called a *Prime Ideal* if $\forall a, b \in R, ab \in P \Rightarrow a \in P$ or $b \in P$.

EXAMPLE 2. In an integral domain D , (0) is a prime ideal. Since $\forall a, b \in D, ab \in (0) \Rightarrow ab = 0 \Rightarrow a = 0$ or $b = 0$ as D is integral domain.

EXAMPLE 3. In $\mathbb{Z}$, the ideal $(3) = \{3n \mid n \in \mathbb{Z}\}$ is prime, since $ab \in (3) \Rightarrow 3 \mid ab \Rightarrow 3 \mid a$ or $3 \mid b \Rightarrow a \in (3)$, or $b \in (3)$. In fact every ideal $(p) = \{pn \mid n \in \mathbb{Z}\}$ where p is a prime number, is a prime ideal of $\mathbb{Z}$.

Theorem 8.4. An ideal P of commutative ring R is prime if and only if R/P is an integral domain.

Proof: Let R/P be an integral domain, then for all $a, b \in R$,
 $ab \in P \Rightarrow \bar{a}\bar{b} = \bar{0} \Rightarrow \bar{a}\bar{b} = \bar{0}$ where $\bar{a} = a + P$
 $\Rightarrow \bar{a} = \bar{0}$ or $\bar{b} = \bar{0}$, as R/P is an integral domain
 $\Rightarrow a \in P$ or $b \in P$.

Thus P is a prime ideal.

Conversely let P be a prime ideal; then

$$\begin{aligned}
 \bar{a}\bar{b} = \bar{0} &\Rightarrow \bar{a}\bar{b} = \bar{0} \\
 &\Rightarrow ab \in P \\
 &\Rightarrow a \in P \text{ or } b \in P \\
 &\Rightarrow \bar{a} = \bar{0} \text{ or } \bar{b} = \bar{0}.
 \end{aligned}$$

Also R/P is commutative as R is commutative.

Consequently R/P is a commutative ring without proper zero divisors. Hence R/P is an integral domain. ■

Definition 8.5. (*Maximal ideal*)

An ideal M of a ring R is said to be a *maximal ideal* of R , if

(i) $M \neq R$,

(ii) there exists no ideal J of R such that $M < J < R$.

Thus from (ii) if $M (\neq R)$ is a maximal ideal then for any ideal J of R , $M \subseteq J \subseteq R$ holds only when either $J=M$ or $J=R$.

EXAMPLE 4. In a division ring D , (0) is a maximal ideal. Trivially $(0) \neq D$ as $1 \in D$ and $1 \neq 0$. Let J be any non-zero ideal of D , then $\exists x (\neq 0)$ in J . But D is a division ring so $x^{-1} \in D$; which gives $1 = x x^{-1} \in J$. Consequently $J=D$. Hence (0) is a maximal ideal.

EXAMPLE 5. In the ring E of even integers, ideal (4) is maximal. As $2 \notin (4)$, $(4) \neq E$. Further let J be an ideal of E such that $(4) < J$. Then there exists an element $x \in J$ such that $x \notin (4)$. In other words x is an even integer not divisible by 4. Consequently $x = 4n + 2$ for some integer n . Now $2 = x - 4n \in J$ as $4n \in (4) < J$ and $x \in J$. Thus every even integral multiple of 2 is in J . Hence $E = J$. This proves the assertion.

Theorem 8.6. An ideal M of a commutative ring R with unity is a maximal ideal if and only if R/M is a field.

Proof: Let M be a maximal ideal of R . Since R is commutative, R/M is commutative. Further 1 is the unity of R gives $\bar{1}$ is the unity of R/M . $M \neq R \Rightarrow \bar{1} \neq \bar{0}$. Thus to show that R/M is a field it is sufficient to prove that every non-zero element of R/M is a unit. Let $\bar{x} (\neq \bar{0}) \in R/M$ then $x \notin M$. Consider $xR = \{xr \mid r \in R\}$. xR is an ideal of R and $x = x1 \in xR$. Since $x \in M + xR$, $x \notin M$, $M < M + xR$. The maximality of M gives $M + xR = R$. Again $1 \in R = M + xR \Rightarrow \exists m \in M, r \in R$ such that $1 = m + xr$. This implies that $1 + M = xr + M \Rightarrow \bar{1} = \bar{x}\bar{r}$. Therefore, $\bar{x}$ is a unit. Hence R/M is a field.

Conversely as the unity of field is different from zero, $\bar{1} \neq \bar{0}$ i.e. $1 \notin M$. Thus $M \neq R$. Let I be any ideal of R such that $M < I$. Choose $a \in I$ such that $a \notin M$. Then $\bar{a} \neq \bar{0}$, so $\bar{a}$ is invertible in R/M . Consequently $\exists \bar{b} \in \bar{R} = R/M$ such that $\bar{a}\bar{b} = \bar{1}$, this yields $1 - ab \in M \Rightarrow 1 - ab \in I$ as $M < I$. This in turn implies that $1 = (1 - ab) + ab \in I$ as $ab \in I$ (I is an ideal and $a \in I$). Hence $I = R$. (Remark following Theorem 7.23). ■

REMARK. If M is any maximal ideal of a commutative ring R with unity then R/M is a field (Theorem 8.6) and we also know that every field is an integral domain, we get M is a prime ideal. Thus every maximal ideal in a commutative ring with unity is prime.

However the converse of this statement is not true.

EXAMPLE 6. Consider the ideal (0) of $\mathbf{Z}$, since $\mathbf{Z}$ is an integral domain, (0) is a prime ideal of $\mathbf{Z}$, but (0) is not maximal since $(0) < (2) < \mathbf{Z}$.

EXAMPLE 7. The ideal (4) in E , the ring of even integers is a maximal ideal (see example 5). But it is not prime since $2 \cdot 2 \in (4)$ but $2 \notin (4)$. Hence in a commutative ring *without unity* a maximal ideal need not be prime.

WORKED-OUT EXERCISES

Exercise 1. Let R be the ring of all real-valued continuous functions on the closed interval $[0, 1]$. Let $M = \{f \in R \mid f(\frac{1}{3}) = 0\}$. show that M is a maximal ideal of R .

Solution. The function $w: [0, 1] \rightarrow \mathbf{R}$ given by $w(x) = 0 \forall x \in [0, 1]$ belongs to M . Hence M is non-empty.

Let $f, g \in M$. Then $(f-g)(\frac{1}{3}) = f(\frac{1}{3}) - g(\frac{1}{3}) = 0 \Rightarrow f-g \in M$. Further let $f \in M, h \in R$ then $hf(\frac{1}{3}) = h(\frac{1}{3})f(\frac{1}{3}) = 0 \Rightarrow hf \in M$. Since R is commutative, $fh = hf \in M$. Hence M is an ideal of R .

Clearly $M \neq R$ as $\theta \in R$ given by $\theta(x) = 1$ does not belong to M .

Let N be an ideal of R such that $M < N$. There exists $\lambda \in N, \lambda \notin M$. This means $\lambda(\frac{1}{3}) \neq 0$. Put $\lambda(\frac{1}{3}) = c$ where $c \neq 0$.

Consider $\mu \in R$ given by $\mu = \lambda - \beta$ where $\beta(x) = c \forall x \in [0, 1]$. Then $\mu(\frac{1}{3}) = \lambda(\frac{1}{3}) - \beta(\frac{1}{3}) = c - c = 0 \Rightarrow \mu \in M \Rightarrow \mu \in N$. Therefore, $\beta = \lambda - \mu \in N$.

Define $\gamma \in R$ by $\gamma(x) = \frac{1}{c} \forall x \in [0, 1]$.

Then $\forall x \in [0, 1], \gamma\beta(x) = \gamma(x)\beta(x) = 1 = \theta(x) \Rightarrow \gamma\beta = \theta \in N$. But θ is unity of R . Hence $N = R$.

Consequently M is a maximal ideal of R .

Exercise 2. For any two ideals A and B of a ring R such that $B \subseteq A$, prove that A/B and B are nilpotent (nil) imply A is nilpotent (nil) ideal.

Solution. Let $\frac{A}{B}$ and B be nilpotent. Then $\left(\frac{A}{B}\right)^n = (B)$.

$B^m = (0)$ for some integers $n, m > 0$. $\left(\frac{A}{B}\right)^n = (B) \Rightarrow A^n \subseteq B \Rightarrow (A^n)^m \subseteq B^m \Rightarrow A^{nm} = (0) \Rightarrow A$ is nilpotent ideal.

Now suppose $\frac{A}{B}$ and B are nil. Let $a \in A$ then $a+B \in \frac{A}{B}$. Since $\frac{A}{B}$ is nil, there exists an integer $n > 0$ such that $(a+B)^n = B \Rightarrow a^n \in B$.

Again as B is nil, there exists an integer $m > 0$ such that $(a^n)^m = 0 \Rightarrow a^{nm} = 0$. This shows that a is nilpotent. Hence A is a nil ideal of R .

PROBLEMS

1. R is a finite non-zero commutative ring. Show that every prime ideal of R is a maximal ideal of R .

2. Give an example of a ring R without unity which has no maximal ideals and in which multiplication is non-trivial. (By trivial multiplication we mean $xy = 0 \forall x, y \in R$).

[Hint. Take $R = \{(a, b) \mid a, b \in \mathbb{Q}\}$. Define $(a, b) + (c, d) = (a + c, b + d)$ and $(a, b)(c, d) = (0, ac)$. Use the fact that $\langle \mathbb{Q}, + \rangle$ has no maximal subgroup.]

3. Prove that in $\mathbb{Z}$, an ideal (n) is maximal if and only if n is prime, where $(n) = \{nx \mid x \in \mathbb{Z}\}$.

4. Prove that an ideal I of a ring R is a maximal ideal if and only if (i) $I \neq R$ and (ii) for each $a \in R$ with $a \notin I$, we have $I + \langle a \rangle = R$.

5. Definition: An ideal M of a ring R is said to be a *minimal ideal* of R if,

(i) $M \neq \{0\}$,

(ii) there exists no ideal J of R such that $\{0\} < J < M$.

Prove that an ideal I of a ring R is a minimal ideal if and only if (i) $I \neq \{0\}$ and (ii) for each $c \neq 0 \in I$, $\langle c \rangle = I$.

6. Let R be the ring of all real-valued continuous functions on the closed interval $[0, 1]$. If M is a maximal ideal of R , prove that there exists a real number γ , $0 \leq \gamma \leq 1$ such that $M = \{f \in R \mid f(\gamma) = 0\}$.

[Hint. Use Heine-Borel's Theorem].

7. Give an example of a non-commutative ring R and an ideal I of R such that (i) $\frac{R}{I}$ is a division ring; (ii) $\frac{R}{I}$ is a field.

[Hint. (ii) $R = \left\{ \begin{pmatrix} a & b \\ 0 & 0 \end{pmatrix} \mid a, b \in \mathbb{Q} \right\}$. Take $I = \left\{ \begin{pmatrix} 0 & x \\ 0 & 0 \end{pmatrix} \mid x \in \mathbb{Q} \right\}$]

8. Let R be a ring. If there exists a positive integer n such that $na = 0$ for all $a \in R$, then the smallest positive integer with this property is called the *characteristic* of the ring R . If no such positive integer n exists i.e. $na = 0 \forall a \in R \Rightarrow n = 0$, then R is said to be *zero characteristic*.

NOTATION. 'Char R ' stands for characteristic of R . Prove that if, n is a positive integer then

(a) The sets $I_n = \{na \mid a \in R\}$ and $J_n = \{a \in R \mid na = 0\}$ are both ideals of R .

(b) $\text{Char}(R/I_n)$ divides n .

(c) If $\text{Char } R \neq 0$ then $\text{Char } R$ divides $n \text{ Char}(R/J_n)$.

2. Homomorphisms

Definition 87. A mapping f of a ring R to a ring R' is called a *ring homomorphism* or simply a *homomorphism* if it satisfies the following properties:

For all $a, b \in R$,

$$(i) f(a+b) = f(a) + f(b).$$

$$(ii) f(ab) = f(a)f(b).$$

REMARK. Clearly (i) shows that a ring homomorphism is a group homomorphism of the additive group $\langle R, + \rangle$ and it preserves the products.

A homomorphism $f: R \rightarrow R'$ is called an *epimorphism* if f is onto. It is called a *monomorphism* if it is 1-1 and *isomorphism* if it is both 1-1 and onto. A homomorphism f of a ring R into itself is called an *endomorphism*. An endomorphism is called an *automorphism* if it is an isomorphism.

EXAMPLE 8. Let S be a subring of R . Define $f: S \rightarrow R$ by $f(x) = x \forall x \in S$. Then clearly for any $x, y \in S$, $f(x+y) = x+y = f(x) + f(y)$ and $f(xy) = xy = f(x)f(y)$. Thus f is a ring homomorphism from S to R . It is called an *inclusion map*.

EXAMPLE 9. Let A be an ideal of a ring R .

Define $f: R \rightarrow R/A$ by $f(r) = r + A \forall r \in R$. f is a ring homomorphism since $f(r+s) = (r+s) + A = (r+A) + (s+A) = f(r) + f(s)$ and $f(rs) = rs + A = (r+A)(s+A) = f(r)f(s)$. Further f is clearly onto. Hence f is an epimorphism.

It is not 1-1 if $A \neq (0)$. To see this take $a (\neq 0)$ in A , then $f(a) = a + A = A$, $f(0) = 0 + A = A$. Hence $f(a) = f(0)$ without a being equal to zero. This proves that f is not 1-1. This map is called *Natural homomorphism*.

EXAMPLE 10. Let $R_1 = \left\{ \begin{pmatrix} a & 0 \\ 0 & 0 \end{pmatrix} \mid a \in R \right\}$ where R is a ring

Define $f: R_1 \rightarrow R$ by $f\left(\begin{pmatrix} a & 0 \\ 0 & 0 \end{pmatrix}\right) = a$ for all $\begin{pmatrix} a & 0 \\ 0 & 0 \end{pmatrix} \in R_1$.

Then f is an isomorphism. This can be verified as under:

$$f\left[\begin{pmatrix} a & 0 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} b & 0 \\ 0 & 0 \end{pmatrix}\right] = f\begin{pmatrix} a+b & 0 \\ 0 & 0 \end{pmatrix} = a+b = f\begin{pmatrix} a & 0 \\ 0 & 0 \end{pmatrix} + f\begin{pmatrix} b & 0 \\ 0 & 0 \end{pmatrix}.$$

$$\text{Also } f\left[\begin{pmatrix} a & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} b & 0 \\ 0 & 0 \end{pmatrix}\right] = f\begin{pmatrix} ab & 0 \\ 0 & 0 \end{pmatrix} = ab = f\begin{pmatrix} a & 0 \\ 0 & 0 \end{pmatrix} f\begin{pmatrix} b & 0 \\ 0 & 0 \end{pmatrix}.$$

Again $f\left(\begin{pmatrix} a & 0 \\ 0 & 0 \end{pmatrix}\right) = f\left(\begin{pmatrix} b & 0 \\ 0 & 0 \end{pmatrix}\right) \Rightarrow a=b \Rightarrow \begin{pmatrix} a & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} b & 0 \\ 0 & 0 \end{pmatrix}$ so f is 1-1.

Finally f is clearly onto. Hence f is an isomorphism.

EXAMPLE 11. Let $M = \{a + b\sqrt{2} \mid a, b \in \mathbb{Z}\}$. Then M is a ring under usual addition and multiplication of irrational numbers.

Define $f: M \rightarrow M$ by $f(a + b\sqrt{2}) = a - b\sqrt{2}$ for all $a + b\sqrt{2} \in M$. f is an automorphism of M (Verify!).

Now we shall establish some theorems on (ring) homomorphisms which are very much analogous to theorems already proved for group homomorphisms. Although proofs of these theorems are necessarily on the same lines as those for the groups, yet we feel it to be desirable from the point of view of clarity to give the proofs of some of the results.

Theorem 8.9. Let $f: R \rightarrow R_1$ be a ring homomorphism. Then

(i) $f(0) = 0$ and (ii) for $\forall a \in R$, $f(-a) = -f(a)$.

Proof: For every $x \in R$, $f(x) = f(x+0) = f(x) + f(0)$.

As $f(x) \in R_1$, we also have $f(x) = f(x) + 0$.

Then $f(x) + f(0) = f(x) + 0$. Hence by cancellation laws with respect to addition, $f(0) = 0$.

Now $a + (-a) = 0 \Rightarrow f[a + (-a)] = f(0) = 0$
 $\Rightarrow f(a) + f(-a) = 0$

Hence $f(-a) = -f(a)$. ■

Corollary 8.10. For all $x, y \in R$, $f(x-y) = f(x) - f(y)$.

Proof: $f(x-y) = f[x + (-y)] = f(x) + f(-y)$
 $= f(x) + [-f(y)]$
 $= f(x) - f(y)$. ■

Definition 8.11. Let $f: R \rightarrow R_1$ be a homomorphism. Kernel of f is defined as the set $\{r \in R \mid f(r) = 0\}$.

We shall denote kernel of f by $\text{Ker } f$.

Theorem 8.12. The kernel of a homomorphism from a ring R to a ring R_1 , is an ideal of R .

Proof: Let $f: R \rightarrow R_1$ be a homomorphism.

As $f(0) = 0$, $0 \in \text{Ker } f$. So $\text{Ker } f$ is non-empty. $x, y \in \text{Ker } f \Rightarrow f(x) = 0$ and $f(y) = 0$.

Thus $f(x-y) = f(x) - f(y) = 0 \Rightarrow x-y \in \text{Ker } f$.

Again $r \in R$, $x \in \text{Ker } f \Rightarrow f(xr) = f(x)f(r) = 0$, since $f(x) = 0$. Similarly $f(rx) = 0$. This implies $xr \in \text{Ker } f$ and $rx \in \text{Ker } f$. Hence $\text{Ker } f$ is an ideal of R . ■

Theorem 8.13. Let $f: R \rightarrow R_1$ be a homomorphism and let A be an ideal of R then $f(A)$ is an ideal of $f(R)$.

Proof: Note that $f(A) = \{x \in R_1 \mid \exists a \in A \text{ such that } x = f(a)\}$.

Clearly $0 \in f(A)$ as $0 = f(0)$ so $f(A)$ is non-empty.

Let $x, y \in f(A)$ then $x=f(a)$, $y=f(b)$ for some $a, b \in A$. So $x-y=f(a)-f(b)=f(a-b) \Rightarrow x-y \in f(A)$ as $a-b \in A$. If $r_1 \in f(R)$ then there exists $r \in R$ such that $r_1=f(r)$. So $r_1x=f(r)f(a)=f(ra) \Rightarrow r_1x \in f(A)$ as $ra \in A$. Similarly $xr_1 \in f(A)$. Hence $f(A)$ is an ideal of $f(R)$. ■

Theorem 8.14. (*Fundamental Theorem of Homomorphism*).

For any ideal A of a ring R , R/A is a homomorphic image of R . Conversely if f is a homomorphism of a ring R onto a ring R_1 , then R_1 is isomorphic to a quotient ring of R . Indeed $R_1 \cong R/\text{Ker } f$.

Proof: Since the natural homomorphism $f: R \rightarrow R/A$ given by $f(x)=x+A \forall x \in R$ is onto, we get R/A is a homomorphic image of R . Conversely let f be a homomorphism of R onto a ring R_1 . Let $A = \text{Ker } f$, then A is an ideal of R . Define $\bar{f}: R/A \rightarrow R_1$ by $\bar{f}(\bar{r})=f(r) \forall \bar{r} = r+A \in R/A$.

Now for any $\bar{r}, \bar{s} \in R/A$, we have

$$\begin{aligned} \bar{r}=\bar{s} &\Leftrightarrow r-s \in A = \text{Ker } f \\ &\Leftrightarrow f(r-s)=0 \\ &\Leftrightarrow f(r)=f(s) \\ &\Leftrightarrow \bar{f}(\bar{r})=\bar{f}(\bar{s}). \end{aligned}$$

This shows that $\bar{f}$ is well-defined as well as 1-1.

$$\text{Also } \bar{f}(\bar{r}+\bar{s})=\bar{f}(\overline{r+s})=f(r+s)=f(r)+f(s)=\bar{f}(\bar{r})+\bar{f}(\bar{s})$$

$$\text{and } \bar{f}(\bar{rs})=\bar{f}(\overline{rs})=f(rs)=f(r)f(s)=\bar{f}(\bar{r})\bar{f}(\bar{s}).$$

This gives that $\bar{f}$ is a monomorphism. Finally as $\bar{f}(\bar{R})=f(R)=R_1$, we get $\bar{f}$ is also onto.

Hence $\bar{f}: R/A \rightarrow R_1$ is an isomorphism and $R/A \cong R_1$ i.e. $R_1 \cong R/A$. ■

Theorem 8.15. Let N be an ideal of R then there is 1-1 correspondence between ideals of R containing N and ideals of R/N .

Proof: Let f be the natural homomorphism from R onto R/N . Let A be an ideal of R containing N , then $f(A)=\{f(a) \mid a \in A\}=\{a+N \mid a \in A\}=A/N$ is an ideal of R/N (Theorem 8.13). Consider any two ideals A and B of R containing N such that $f(A)=f(B)$ i.e. $A/N=B/N$. Then for any $a \in A \exists b \in B$ such that $a+N=b+N$. This implies $a-b \in N \subseteq B \Rightarrow a=(a-b)+b \in B$. Consequently $A \subseteq B$. Similarly it can be shown that $B \subseteq A$. So we get $A=B$. Hence the mapping $A \rightarrow f(A)$ of the set $\mathcal{F}$ of all ideals of R containing N to the set $\bar{\mathcal{F}}$ of all ideals of R/N is 1-1. To show that it is onto we show that any ideal of R/N is of the form $f(A)$ for some $A \in \mathcal{F}$. Let X be any ideal of R/N . Define $A=\{x \in R \mid f(x) \in X\}$. Since for each $n \in N$, $f(n)=\bar{0} \in X$, we get $n \in A$ i.e. $N \subseteq A$. Evidently each member of X is of the form $f(a)$, $a \in A$. If we show that A is an ideal of R we shall get $X=f(A)$ with $A \in \mathcal{F}$ and that is turn will give $A \leftrightarrow f(A)$ is a 1-1

correspondence between F and $\overline{F}$. Let $a, b \in A$, $r \in R$, then $f(a), f(b) \in X$, $f(r) \in R/N$. Thus $f(a-b) = f(a) - f(b) \in X$, $f(ar) = f(a)f(r) \in X$ and $f(ra) = f(r)f(a) \in X$, as X is an ideal of R/N . As a consequence $a-b \in A$, $ar \in A$, $ra \in A$ and A is an ideal of R . So $A \in F$. This proves the theorem. ■

Following theorem gives an apparent generalization of the above theorem. The proof is exactly on same lines, only one has to replace f by g and take $N = \text{Ker } g$.

Theorem 8.16. *If g is a homomorphism of a ring R onto a ring R_1 , then there exists a 1-1 correspondence between the family of ideals of R containing $\text{Ker } g$ and the family of all ideals of R_1 , given by $A \leftrightarrow g(A)$. ■*

Theorem 8.17. (First Theorem of Isomorphism).

Let $B \subseteq A$ be two ideals of a ring R then $R/A \cong (R/B)/(A/B)$.

Proof: Define $f: R/B \rightarrow R/A$ by $f(r+B) = r+A \forall r \in R$. f is well-defined, since $r+B = s+B \Rightarrow r-s \in B \subseteq A \Rightarrow r+A = s+A$.

We leave it for the readers to show that f is an epimorphism with A/B as its kernel. ■

Theorem 8.18. (Second Theorem of Isomorphism).

Let A and B be two ideals of a ring R , then $\frac{A+B}{A} \cong \frac{B}{A \cap B}$.

Proof: Define $f: B \rightarrow \frac{A+B}{A}$ by $f(b) = \overline{b} \forall b \in B$.

Let $b_1, b_2 \in B$, then

(i) $f(b_1+b_2) = \overline{b_1+b_2} = \overline{b_1} + \overline{b_2} = f(b_1) + f(b_2)$.

(ii) $f(b_1 b_2) = \overline{b_1 b_2} = \overline{b_1} \overline{b_2} = f(b_1) f(b_2)$. Thus f is a homomorphism.

Now $\overline{x} \in \frac{A+B}{A} \Rightarrow x = a+b$ for some $a \in A, b \in B$.

So $\overline{x} = a+b+A = b+a+A = b+A$ as $a \in A$ i.e. $\overline{x} = \overline{b} = f(b)$, hence f is onto.

Finally $\text{Ker } f = A \cap B$, since

$$b \in \text{Ker } f \Leftrightarrow \overline{b} = \overline{0}$$

$$\Leftrightarrow b \in A$$

$$\Leftrightarrow b \in A \cap B \text{ as } b \in B.$$

Hence by the Fundamental Theorem of Homomorphism

$$\frac{B}{A \cap B} \cong \frac{A+B}{A} \text{ i.e. } \frac{A+B}{A} \cong \frac{B}{A \cap B}. \blacksquare$$

PROBLEMS

1. Let R be a commutative ring such that $px=0 \forall x \in R$, p being a prime number. Define $f: R \rightarrow R$ such that $f(x) = x^p$. Show that f is a

homomorphism.

2. Let f be a homomorphism of a ring R onto R' with $\text{Ker } f = N$. Prove each of the following results.

(i) If A is a right (left) ideal in R , then $f(A)$ is a right (left) ideal in R' .

(ii) If B is a right (left) ideal in R' , then $f^{-1}(B) = \{r \in R \mid f(r) \in B\}$ is a right (left) ideal of R , which contains N .

(iii) If A and B are ideals of R such that $N \subseteq A$ and $N \subseteq B$, then $A \subseteq B$ if and only if $f(A) \subseteq f(B)$.

3. If f is a homomorphism from a ring R into a ring S and g is a homomorphism from S into a ring T , show that gf is a homomorphism from R into T .

4. Complete the proof of Theorem 8.17.

5. Let A and B be two subrings of a ring R such that for every $a \in A$, $b \in B$, ab and ba both are in A . Prove the following.

(i) $A+B$ is a subring of R .

(ii) A and $A \cap B$ are ideals of $A+B$ and B respectively.

(iii) $\frac{A+B}{A} \cong \frac{B}{A \cap B}$.

How does this result generalize the Second Theorem of Isomorphism. Give an example to show that sum of two subrings may not be a subring.

[Hint Take $S = \left\{ \begin{pmatrix} 0 & a \\ 0 & 0 \end{pmatrix} \mid a \in \mathbb{Z} \right\}$ and $T = \left\{ \begin{pmatrix} 0 & 0 \\ b & 0 \end{pmatrix} \mid b \in \mathbb{Z} \right\}$

S and T are subrings of the ring $\mathbb{Z}_2$ of all 2×2 matrices over $\mathbb{Z}$. But $S+T$ is not a subring of $\mathbb{Z}_2$.]

6. A homomorphism f from a ring R to a ring S is said to be a *zero homomorphism* if $f(R) = 0$. Prove that a homomorphism from a field to any ring is either a monomorphism or a zero homomorphism.

7. If $f: R \rightarrow R_1$ is a homomorphism and R has unity 1, then prove that $f(1)$ is the unity of $f(R)$.

8. Let $f: R \rightarrow S$ be an epimorphism. If A and B are ideals in R , U, V are ideals in S , prove that

(i) $f(A+B) = f(A) + f(B)$.

(ii) $f(AB) = f(A)f(B)$.

(iii) $f^{-1}(U+V) = f^{-1}(U) + f^{-1}(V)$.

(iv) $f^{-1}(UV) \supseteq f^{-1}(U)f^{-1}(V)$.

Give an example to show that in general $f^{-1}(UV) \neq f^{-1}(U)f^{-1}(V)$.

[Hint. Let R be an additive Abelian group with more than one element. Define $xy = 0 \forall x, y \in R$. Then R is a ring. Take $S = \{0\}$ and define $f: R \rightarrow S$ by $f(x) = 0 \forall x \in R$. Take $U = V = \{0\}$ then UV

$=\{0\}$ and $f^{-1}(UV)=R$, while $f^{-1}(U)f^{-1}(V)=RR=\{0\}$.]

9. Show that $f: \mathbf{C} \rightarrow \mathbf{R}_2$ given by $f(a+bi) = \begin{pmatrix} a & b \\ -b & a \end{pmatrix}$ is a monomorphism, where $\mathbf{R}_2$ is the ring of all 2×2 matrices over $\mathbf{R}$.

10. Let $(p) = \{pn \mid n \in \mathbf{Z}, p \neq 0\}$, show that $\mathbf{Z}/(p)$ is a field if and only if p is a prime number.

[Hint. Any non-zero finite integral domain is a field.]

11. (Fermat). Let p be a prime number. Prove that for all $a \in \mathbf{Z}$ not divisible by p , $a^{p-1} \equiv 1 \pmod{p}$.

[Hint. $\mathbf{Z}/(p)$ is a field, so $\bar{a} \neq \bar{0}$ and non-zero elements of a field form a group under multiplication.]

12. If p is an odd prime number then show that in $\mathbf{Z}/(p)$ there are exactly two elements $\bar{x}$ such that $\bar{x}^2 = \bar{1}$. Hence deduce Wilson's Theorem. "If p is a prime number then $(p-1)! \equiv -1 \pmod{p}$."

13. A is an ideal of a commutative ring R . Let $\sqrt{A}$ be defined as in Section 4, problem 6. Show that $R/\sqrt{A}$ has no non-zero nilpotent elements.

14. In a ring R define S as the set $\{ab-ba \mid a, b \in R\}$. Show that for any ideal A of R , the quotient ring R/A is commutative if and only if $S \subseteq A$.

15. An ideal Q of a commutative ring R , is called a *primary ideal* if (i) $Q \neq R$ and (ii) for all $a, b \in R$, $ab \in Q$ and $a \notin Q \Rightarrow b^n \in Q$ for some positive integer n . Show that Q is a primary ideal of R if and only if every zero-divisor of R/Q is nilpotent.

16. An ideal A of a commutative ring R is called a *semi-prime ideal* if for $\forall a \in R$, $a^n \in A \Rightarrow a \in A$. Show that A is semi-prime ideal of R if and only if R/A has non-zero nilpotent elements.

17. I is an ideal of a ring R , prove that the matrix ring $(R/I)_n$ is isomorphic to R_n/I_n .

[Hint. Define $f: R_n \rightarrow (R/I)_n$ by $f(a_{ij}) = (a_{ij} + I) \forall (a_{ij}) \in R_n$.]

3. Field of Quotients and Embedding Theorems

Some types of rings are easier to study and their structures are better known than those of the other types. If one can embed a certain ring R in a ring S , such that the structure of the latter is better known than that of former, then by using the properties of S one can say a lot about the properties of R . The process of embedding of one ring into another has been quite fruitfully used for the development of ring theory. There are many procedures of embedding of one ring into another. Here in this section we confine ourselves to only three methods: (I) Embedding of a ring in a ring with unity, (II) Embedding of a domain in a field, and (III) Embedding of a ring in a

ring of endomorphisms of some Abelian group.

(I) *Embedding of a ring in a ring with unity.*

Let R be a ring, consider $R \times \mathbf{Z} = \{(r, n) \mid r \in R, n \in \mathbf{Z}\}$. $R \times \mathbf{Z}$ can be made into a ring by defining addition and multiplication as under:

For all $(r, n), (s, m) \in R \times \mathbf{Z}$, $(r, n) + (s, m) = (r + s, n + m)$
and $(r, n)(s, m) = (rs + ns + mr, nm)$.

It can be checked that $R \times \mathbf{Z}$ is a ring with unity $(0, 1)$.

Theorem 8.19. Every ring can be embedded in a ring with unity.

Proof: Let R be a ring, then as remarked above $R_1 = R \times \mathbf{Z} = \{(r, n) \mid r \in R, n \in \mathbf{Z}\}$ is a ring. We define $f: R \rightarrow R_1$ by $f(r) = (r, 0) \forall r \in R$. Clearly f is a homomorphism. f is also 1-1 as $f(r) = f(s) \Rightarrow (r, 0) = (s, 0) \Rightarrow r = s$.

Hence $R \cong f(R) \subseteq R_1$. Consequently R is embeddable in R_1 which has unity namely $(0, 1)$. ■

REMARK. The embedding discussed in the above theorem is *very nice* in the sense that $f(R)$ is an ideal of R_1 . To check this, note that for $(r, 0) \in f(R)$, $(s, n) \in R_1$, $(r, 0)(s, n) = (rs + 0s + nr, 0n) = (rs + nr, 0) \in f(R)$. Similarly $(s, n)(r, 0) = (sr + nr, 0) \in f(R)$. Thus R , under identification of r with $(r, 0)$, becomes an ideal of R_1 .

(II) *Embedding of a domain in a field.*

Every subring of a field is an integral domain. Naturally one can ask that given an integral domain, can we find a field which has that domain, as its subring? Or formally speaking, can an integral domain be embedded in a field? The question is of vital importance when one considers the solution of linear equation $ax = b$, $a \neq 0$, in an integral domain D . Of course if a solution exists in D , it must be unique (Why?). It is quite possible that no solution may exist in D . For example in $\mathbf{Z}$, $3x = 5$ has no solution! But in a field F , $ax = b$, $a \neq 0$ always has a solution namely $a^{-1}b$. Thus if a domain D can be embedded in a field F then an equation $ax = b$, $a \neq 0$ with coefficients in D will at least have a solution in F . Now we endeavour to construct a field in which a given integral domain can be embedded. Our procedure is motivated by the method of construction of rational numbers from integers.

Let D be an integral domain with at least two elements and let $D_0 = D - \{0\}$. Consider $D \times D_0 = \{(a, b) \mid a, b \in D, b \neq 0\}$.

Define a relation ' $\sim$ ' on $D \times D_0$ as under:

For all $(a, b), (c, d) \in D \times D_0$, $(a, b) \sim (c, d)$ if and only if $ad = bc$.

Lemma 8.20. ' $\sim$ ' is an equivalence relation on $D \times D_0$.

Proof: Let $(a, b), (c, d), (e, f) \in D \times D_0$.

(i) *Reflexivity:* As $ab = ba$, putting $d = b$ and $c = a$ in the definition of ' $\sim$ ' we see that $(a, b) \sim (a, b)$.

(ii) Symmetry: $(a,b) \sim (c,d) \Rightarrow ad=bc$

$$\Rightarrow cb=da \Rightarrow (c,d) \sim (a,b).$$

(iii) Transitivity: $(a,b) \sim (c,d)$ and $(c,d) \sim (e,f)$

$$\Rightarrow ad=bc \text{ and } cf=de,$$

$$\Rightarrow adf=bcf$$

$$\Rightarrow adf=bde \text{ as } cf=de,$$

$$\Rightarrow af=be \text{ as } d \neq 0 \text{ and cancellation law holds in } D_0.$$

$$\Rightarrow (a,b) \sim (e,f).$$

Hence ' $\sim$ ' is an equivalence relation on $D \times D_0$. ■

REMARK. From the definition of equivalence relation one sees immediately that for any non-zero elements a, b and c of D , $(0,a) \sim (0,b)$; $(a,a) \sim (b,b)$ and $(ac, bc) \sim (a,b)$.

Let a/b be the equivalence class of (a,b) . By properties of equivalence classes it is evident that $a/b = c/d$ if and only if $ad=bc$. Now if F is the family of all equivalence classes a/b of $D \times D_0$, we shall show that under suitably defined addition and multiplication, F becomes a field. Once again the addition and multiplication of rational numbers will give us a clue. So we define for all $a/b, c/d \in F$, $a/b + c/d = (ad+bc)/bd$ and $(a/b)(c/d) = ac/bd$. Note that as D is an integral domain, $b \neq 0, d \neq 0 \Rightarrow bd \neq 0$, so $(ad+bc)/bd$ and $ac/bd \in F$. In order that our definitions should make sense we must show that they are independent of the choice of representatives of equivalence classes involved in the operations. In other words we must show that if $a/b = a'/b'$ and $c/d = c'/d'$ then $(ad+bc)/bd = (a'd'+b'c')/b'd'$ and $ac/bd = a'c'/b'd'$.

Lemma 8.21. *The addition and multiplication as defined above are well-defined.*

Proof: Let $a/b = a'/b'$ and $c/d = c'/d'$

$$\text{then } ab' = a'b \text{ and } cd' = c'd.$$

$$\text{Now (1) } \Rightarrow ab'dd' = a'bdd' \text{ and } bb'cd' = bb'c'd' \quad \dots (1)$$

$$\Rightarrow ab'dd' + bb'cd' = a'bdd' + bb'c'd'$$

$$\Rightarrow (ad+bc)b'd' = (a'd'+b'c')bd$$

$$\Rightarrow \frac{ad+bc}{bd} = \frac{a'd'+b'c'}{b'd'}. \text{ (Note. } \frac{x}{y} \text{ stands for } x/y)$$

Hence addition is well-defined.

$$\text{Finally (1) } \Rightarrow ab'cd' = a'bc'd'$$

$$\Rightarrow (ac)(b'd') = (ba)(d'c')$$

$$\Rightarrow \frac{ac}{bd} = \frac{a'c'}{b'd'}.$$

Consequently multiplication is also well-defined. Hence the lemma follows. ■

Theorem 8.22. *$\langle F, +, \cdot \rangle$ is a field and D can be embedded into F .*

Proof: From the remark following Lemma 8.20, it follows that for $a \neq 0 \in D$, $\frac{0}{a}$ and $\frac{a}{a} \in F$ and $\frac{0}{a} \neq \frac{a}{a}$. So F has at least two elements. Further notice that for any two non-zero elements a, b in D , $\frac{0}{a} = \frac{0}{b}$; also if $c \neq 0$ then $\frac{ac}{bc} = \frac{a}{b}$.

(1) $+$ is *Associative*: For all $\frac{a}{b}, \frac{c}{d}, \frac{e}{f} \in F$

$$\begin{aligned} \left(\frac{a}{b} + \frac{c}{d} \right) + \frac{e}{f} &= \frac{ad+bc}{bd} + \frac{e}{f} \\ &= \frac{(ad+bc)f + (bd)e}{(bd)f} \\ &= \frac{a(df) + (cf+de)b}{b(df)} \\ &= \frac{a}{b} + \frac{cf+de}{df} \\ &= \frac{a}{b} + \left(\frac{c}{d} + \frac{e}{f} \right). \end{aligned}$$

(2) $+$ is *commutative*: For all $\frac{a}{b}, \frac{c}{d} \in F$

$$\begin{aligned} \frac{a}{b} + \frac{c}{d} &= \frac{ad+bc}{bd} = \frac{bc+ad}{db} \\ &= \frac{c}{d} + \frac{a}{b}. \end{aligned}$$

(3) *Existence of zero element*:

Now for $u \neq 0$ in D , $\frac{0}{u} \in F$ is such that $\frac{a}{b} + \frac{0}{u} = \frac{au}{bu} = \frac{a}{b}$.

Hence $\frac{0}{u}$ is zero element of F . Let us denote it by 0.

(4) *Existence of additive inverse*:

$$\text{Since } \frac{a}{b} + \left(\frac{-a}{b} \right) = \frac{ab-ba}{b^2} = \frac{0}{b^2}$$

$$\text{and } \frac{0}{u} = \frac{0}{b^2}, \text{ we have } \frac{a}{b} + \left(\frac{-a}{b} \right) = 0.$$

$$\text{Thus } -\left(\frac{a}{b} \right) = \frac{(-a)}{b} \in F.$$

(5) *Associativity of multiplication*:

For all $\frac{a}{b}, \frac{c}{d}, \frac{e}{f} \in F$

$$\left(\frac{a}{b} \cdot \frac{c}{d} \right) \left(\frac{e}{f} \right) = \left(\frac{ac}{bd} \right) \left(\frac{e}{f} \right) = \frac{(ac)e}{(bd)f}$$

(ii) Symmetry: $(a,b) \sim (c,d) \Rightarrow ad=bc$

$$\Rightarrow cb=da \Rightarrow (c,d) \sim (a,b).$$

(iii) Transitivity: $(a,b) \sim (c,d)$ and $(c,d) \sim (e,f)$

$$\Rightarrow ad=bc \text{ and } cf=de,$$

$$\Rightarrow adf=bcf$$

$$\Rightarrow adf=bde \text{ as } cf=de,$$

$$\Rightarrow af=be \text{ as } d \neq 0 \text{ and cancellation law holds in } D_0.$$

$$\Rightarrow (a,b) \sim (e,f).$$

Hence ' $\sim$ ' is an equivalence relation on $D \times D_0$. ■

REMARK. From the definition of equivalence relation one sees immediately that for any non-zero elements a, b and c of D , $(0, a) \sim (0, b)$; $(a, a) \sim (b, b)$ and $(ac, bc) \sim (a, b)$.

Let a/b be the equivalence class of (a, b) . By properties of equivalence classes it is evident that $a/b = c/d$ if and only if $ad=bc$. Now if F is the family of all equivalence classes a/b of $D \times D_0$, we shall show that under suitably defined addition and multiplication, F becomes a field. Once again the addition and multiplication of rational numbers will give us a clue. So we define for all $a/b, c/d \in F$, $a/b + c/d = (ad+bc)/bd$ and $(a/b)(c/d) = ac/bd$. Note that as D is an integral domain, $b \neq 0, d \neq 0 \Rightarrow bd \neq 0$, so $(ad+bc)/bd$ and $ac/bd \in F$. In order that our definitions should make sense we must show that they are independent of the choice of representatives of equivalence classes involved in the operations. In other words we must show that if $a/b = a'/b'$ and $c/d = c'/d'$ then $(ad+bc)/bd = (a'd' + b'c')/b'd'$ and $ac/bd = a'c'/b'd'$.

Lemma 8.21. *The addition and multiplication as defined above are well-defined.*

Proof: Let $a/b = a'/b'$ and $c/d = c'/d'$

$$\text{then } ab' = a'b \text{ and } cd' = c'd.$$

$$\text{Now (1) } \Rightarrow ab'dd' = a'bdd' \text{ and } bb'cd' = bb'c'd$$

$$\Rightarrow ab'dd' + bb'cd' = a'bdd' + bb'c'd$$

$$\Rightarrow (ad+bc)b'd' = (a'd' + b'c')bd$$

$$\Rightarrow \frac{ad+bc}{bd} = \frac{a'd' + b'c'}{b'd'}. \quad (\text{Note. } \frac{x}{y} \text{ stands for } x/y)$$

Hence addition is well-defined.

$$\text{Finally (1) } \Rightarrow ab'cd' = a'bc'd$$

$$\Rightarrow (ac)(b'd') = (ba)(d'c')$$

$$\Rightarrow \frac{ac}{bd} = \frac{a'c'}{b'd'}.$$

Consequently multiplication is also well-defined. Hence the lemma follows. ■

Theorem 8.22. *$\langle F, +, \cdot \rangle$ is a field and D can be embedded into F .*

Proof: From the remark following Lemma 8.20, it follows that for $a \neq 0 \in D$, $\frac{0}{a}$ and $\frac{a}{a} \in F$ and $\frac{0}{a} \neq \frac{a}{a}$. So F has at least two elements. Further notice that for any two non-zero elements a, b in D , $\frac{0}{a} = \frac{0}{b}$; also if $c \neq 0$ then $\frac{ac}{bc} = \frac{a}{b}$.

(1) $+$ is *Associative*: For all $\frac{a}{b}, \frac{c}{d}, \frac{e}{f} \in F$

$$\begin{aligned} \left(\frac{a}{b} + \frac{c}{d} \right) + \frac{e}{f} &= \frac{ad+bc}{bd} + \frac{e}{f} \\ &= \frac{(ad+bc)f + (bd)e}{(bd)f} \\ &= \frac{a(df) + (cf+de)b}{b(df)} \\ &= \frac{a}{b} + \frac{cf+de}{df} \\ &= \frac{a}{b} + \left(\frac{c}{d} + \frac{e}{f} \right). \end{aligned}$$

(2) $+$ is *commutative*: For all $\frac{a}{b}, \frac{c}{d} \in F$

$$\begin{aligned} \frac{a}{b} + \frac{c}{d} &= \frac{ad+bc}{bd} = \frac{bc+ad}{db} \\ &= \frac{c}{d} + \frac{a}{b}. \end{aligned}$$

(3) *Existence of zero element*:

Now for $u \neq 0$ in D , $\frac{0}{u} \in F$ is such that $\frac{a}{b} + \frac{0}{u} = \frac{au}{bu} = \frac{a}{b}$.

Hence $\frac{0}{u}$ is zero element of F . Let us denote it by 0.

(4) *Existence of additive inverse*:

$$\text{Since } \frac{a}{b} + \left(\frac{-a}{b} \right) = \frac{ab-ba}{b^2} = \frac{0}{b^2}$$

$$\text{and } \frac{0}{u} = \frac{0}{b^2}, \text{ we have } \frac{a}{b} + \left(\frac{-a}{b} \right) = 0.$$

$$\text{Thus } -\left(\frac{a}{b} \right) = \frac{(-a)}{b} \in F.$$

(5) *Associativity of multiplication*:

For all $\frac{a}{b}, \frac{c}{d}, \frac{e}{f} \in F$

$$\left(\frac{a}{b} \cdot \frac{c}{d} \right) \left(\frac{e}{f} \right) = \left(\frac{ac}{bd} \right) \left(\frac{e}{f} \right) = \frac{(ac)e}{(bd)f}$$

$$= \frac{a(ce)}{b(df)} = \frac{a}{b} \left(\frac{c}{d} \frac{e}{f} \right).$$

(6) *Commutativity of multiplication:*

$$\text{For all } \frac{a}{b}, \frac{c}{d} \in F, \frac{a}{b} \frac{c}{d} = \frac{ac}{bd} = \frac{ca}{ab} = \frac{c}{d} \frac{a}{b}.$$

(7) *Existence of unity:* Now if $u \neq 0 \in D$ then $\frac{u}{u} \in F$ is such that

$$\frac{a}{b} \frac{u}{u} = \frac{au}{bu} = \frac{a}{b}. \text{ This gives that } \frac{u}{u} \text{ is multiplicative identity i.e. unity for } F. \text{ We shall denote it by } 1.$$

(8) *Existence of multiplicative inverses of non-zero elements:*

$$\text{Now } \frac{a}{b} \neq 0 \Rightarrow a \neq 0, b \neq 0 \Rightarrow \frac{b}{a} \in F$$

$$\text{and } \frac{a}{b} \frac{b}{a} = \frac{ab}{ab} = \frac{u}{u} = 1.$$

$$\text{Thus } \left(\frac{a}{b} \right)^{-1} = \frac{b}{a}.$$

(9) *Distributivity:* For all $\frac{a}{b}, \frac{c}{d}, \frac{e}{f} \in F$

$$\begin{aligned} \frac{a}{b} \left(\frac{c}{d} + \frac{e}{f} \right) &= \frac{a}{b} \left(\frac{cf+de}{df} \right) \\ &= \frac{a(cf+de)}{b(df)} \end{aligned}$$

$$\begin{aligned} \text{Also } \frac{a}{b} \frac{c}{d} + \frac{a}{b} \frac{e}{f} &= \frac{ac}{bd} + \frac{ae}{bf} \\ &= \frac{(ac)bf + (ae)bd}{(bd)bf} \\ &= \frac{[(ac)f + (ae)d]b}{(bdf)b} \\ &= \frac{a(cf+de)}{b(df)}. \end{aligned}$$

Hence multiplication is distributive over addition. Consequently $\langle F, +, \cdot \rangle$ is a field.

Finally define $f: D \rightarrow F$ as under:

$$\text{Take } a(\neq 0) \in D, \text{ then } \forall x \in D \text{ put } f(x) = \frac{xa}{a}.$$

$$\begin{aligned} \text{Let } x, y \in D, \text{ then } f(x) = f(y) &\Rightarrow \frac{xa}{a} = \frac{ya}{a} \\ &\Rightarrow xa^2 = ya^2 \Rightarrow x = y. \end{aligned}$$

Thus f is 1-1.

$$\text{Also } f(x+y) = \frac{(x+y)a}{a} = \frac{(x+y)a^2}{a^2} = \frac{xa}{a} + \frac{ya}{a}$$

$$=f(x)+f(y).$$

$$\text{and } f(xy) = \frac{xya}{a} = \frac{(xa)(ya)}{a^2} = \frac{xa}{a} \frac{ya}{a} = f(x)f(y).$$

This proves that f is a monomorphism of D into F . Hence D is embeddable in a field F . ■

Definition 8.23. Let D be an integral domain with more than one element. A *field of quotients* of D is a pair (F, σ) where F is a field and σ is a monomorphism of D into F such that every $z \in F$ is expressible as $\sigma(x)/\sigma(y)$ for some $x, y \in D$ with $y \neq 0$.

REMARK. When there is no confusion regarding the monomorphism σ then F itself will be referred to as the field of quotients of D . Now as $D \cong \sigma(D) \subseteq F$, we shall normally identify D with its image $\sigma(D)$ and regard D itself as a subring of F . In that case each $a \in D$ is identified with $\sigma(a)$ and hence, then each $z \in F$ is of the form x/y ; $x, y \in D$ with $y \neq 0$.

Theorem 8.24. Let D be an integral domain containing at least two elements then D has a field of quotients. If (F_1, σ_1) and (F_2, σ_2) are two fields of quotients then there exists an isomorphism η of F_1 onto F_2 such that $\eta \sigma_1 = \sigma_2$.

Proof: In Theorem 8.22 we constructed a field F such that $f: D \rightarrow F$ with $f(x) = xa/a \forall x \in D$ and a , a non-zero fixed element of D . Also we saw that f was a monomorphism. Now each $z \in F$ is of the form x/y , $x, y \in D$, $y \neq 0$. Then $z = (xa/a)(a/ya) = f(x)/f(y)$. Hence (F, f) is a field of quotients of D .

Let (F_1, σ_1) and (F_2, σ_2) be two fields of quotients of D . Define $\eta: F_1 \rightarrow F_2$ as under:

Take $z \in F_1$, then by definition $z = \frac{\sigma_1(x)}{\sigma_1(y)}$ for some $x, y \in D$, $y \neq 0$.

$$\text{Put } \eta(z) = \frac{\sigma_2(x)}{\sigma_2(y)}.$$

η is well-defined since if $z = \frac{\sigma_1(x)}{\sigma_1(y)} = \frac{\sigma_1(u)}{\sigma_1(v)}$ for some $x, y, u, v \in D$,

with $y \neq 0$ and $v \neq 0$, we get $\sigma_1(xv) = \sigma_1(uy)$

$$\Rightarrow xv = uy \text{ as } \sigma_1 \text{ is } 1-1 \Rightarrow \sigma_2(xv) = \sigma_2(uy)$$

$$\Rightarrow \sigma_2(x) \sigma_2(v) = \sigma_2(u) \sigma_2(y) \Rightarrow \frac{\sigma_2(x)}{\sigma_2(y)} = \frac{\sigma_2(u)}{\sigma_2(v)}.$$

Hence η is well-defined.

Now for $z = \frac{\sigma_1(x)}{\sigma_1(y)}$, $t = \frac{\sigma_1(u)}{\sigma_1(v)} \in F$

$$\eta(z+t) = \eta \left[\frac{\sigma_1(x)}{\sigma_1(y)} + \frac{\sigma_1(u)}{\sigma_1(v)} \right] = \eta \left[\frac{\sigma_1(xv + yu)}{\sigma_1(yv)} \right]$$

$$\begin{aligned} &= \frac{\sigma_2(xv + yu)}{\sigma_2(yv)} = \frac{\sigma_2(x)}{\sigma_2(y)} + \frac{\sigma_2(u)}{\sigma_2(y)} \\ &= \eta(z) + \eta(t) \end{aligned}$$

$$\begin{aligned} \text{Also } \eta(zt) &= \eta \left[\frac{\sigma_1(x)}{\sigma_1(y)} \cdot \frac{\sigma_1(u)}{\sigma_1(v)} \right] = \eta \left[\frac{\sigma_1(xu)}{\sigma_1(yv)} \right] = \frac{\sigma_2(xu)}{\sigma_2(yv)} \\ &= \frac{\sigma_2(x)}{\sigma_2(y)} \cdot \frac{\sigma_2(u)}{\sigma_2(v)} = \eta(z) \eta(t). \end{aligned}$$

This proves that η is a homomorphism. Now we show that η is onto. Let $z' \in F_2$, then by definition $z' = \frac{\sigma_2(x')}{\sigma_2(y')}$ for some $x', y' \in D$ with $y' \neq 0$. Clearly $\frac{\sigma_1(x')}{\sigma_1(y')} \in F_1$ is such that $\eta \left[\frac{\sigma_1(x')}{\sigma_1(y')} \right] = z'$. This shows that η is onto.

$$\begin{aligned} \text{Finally } z \in \text{Ker } \eta &\Rightarrow \eta(z) = 0 \Rightarrow \frac{\sigma_2(x)}{\sigma_2(y)} = 0 \text{ where } z = \frac{\sigma_1(x)}{\sigma_1(y)}. \\ &\Rightarrow \sigma_2(x) = 0 \Rightarrow x = 0 \text{ as } \sigma_2 \text{ is } 1-1 \\ &\Rightarrow \sigma_1(x) = 0 \Rightarrow z = \frac{\sigma_1(x)}{\sigma_1(y)} = 0. \end{aligned}$$

So that $\text{Ker } \eta = (0)$ and η is 1-1 mapping. Hence η is an isomorphism of F_1 onto F_2 .

As by definition of η , for any $x \in D$, $\eta[\sigma_1(x)] = \sigma_2(x)$ we get $\eta\sigma_1 = \sigma_2$. This completes the proof. ■

Theorem 8.25. If a field K contains a non-zero integral domain D then K contains a field of quotients of D .

Proof: Let $F = \left\{ \frac{a}{b} \mid a, b \in D, b \neq 0 \right\}$. (Note that by $\frac{a}{b}$ we mean

ab^{-1}). Since D is non-zero, $\exists a (\neq 0) \in D$. Then $1 = \frac{a}{a} \in F$. Now for

any $x = \frac{a}{b}$, $y = \frac{c}{d}$, $x - y = \frac{ad - bc}{bd} \in F$ and $xy = \frac{ac}{bd} \in F$. Thus F

is a subring of K containing 1. Now if $x \neq 0$ then $a \neq 0$ and $x' = \frac{b}{a}$

$\in F$ such that $xx' = \frac{a}{b} \cdot \frac{b}{a} = 1$. Hence $x^{-1} = x' \in F$. This proves that

F is a field.

Given $z (\neq 0) \in D$ we can write $z = za/a$ (under identification, as remarked earlier) and $(za)/a \in F$. Thus we have $D \subseteq F$. From the definition it is clear that F is the field of quotients of D contained in K . This proves the theorem. ■

(III) Embedding of a ring in a ring of endomorphism of some Abelian group.

Let V be an additive Abelian group. We denote by $\text{Hom}(V, V)$ or b_V

End V , the set of all endomorphisms of V . First of all we show that $\text{Hom}(V, V)$ is a ring under suitably defined addition and multiplication. We define addition and multiplication on $\text{Hom}(V, V)$ as under. For all $f, g \in \text{Hom}(V, V)$, $f+g$ is the mapping from V to V given by $(f+g)(v) = f(v) + g(v) \forall v \in V$ and fg is the composite mapping of V into V i.e. $fg(v) = f[g(v)] \forall v \in V$.

Theorem 8.26. *$\text{Hom}(V, V)$ is a ring with unity.*

Proof: We define addition and multiplication as discussed above. First of all we show that for all $f, g \in \text{Hom}(V, V)$, $f+g$ and fg both are in $\text{Hom}(V, V)$.

$$\begin{aligned} \text{For } u, v \in V, (f+g)(u+v) &= f(u+v) + g(u+v) \\ &= [f(u) + f(v)] + [g(u) + g(v)] \\ &= [f(u) + g(u)] + [f(v) + g(v)] \text{ as } V \text{ is addi-} \\ &\quad \text{tive Abelian group.} \\ &= (f+g)(u) + (f+g)(v) \end{aligned}$$

Hence $f+g \in \text{Hom}(V, V)$.

$$\begin{aligned} \text{Further } (fg)(u+v) &= f[g(u+v)] \\ &= f[g(u) + g(v)] \\ &= fg(u) + fg(v). \end{aligned}$$

Hence $fg \in \text{Hom}(V, V)$.

(1) $+$ is Associative: For all $f, g, h \in \text{Hom}(V, V)$

$$\begin{aligned} [(f+g)+h](v) &= (f+g)(v) + h(v) \forall v \in V \\ &= [f(v) + g(v)] + h(v) \\ &= f(v) + [g(v) + h(v)] \\ &= f(v) + (g+h)(v) \\ &= [f+(g+h)](v). \end{aligned}$$

This implies that $(f+g)+h = f+(g+h)$.

(2) $+$ is Commutative: For all $f, g \in \text{Hom}(V, V)$

$$\begin{aligned} (f+g)(v) &= f(v) + g(v) \forall v \in V \\ &= g(v) + f(v) \\ &= (g+f)(v) \end{aligned}$$

This implies that $f+g = g+f$.

(3) Existence of zero element: Clearly the mapping $O: V \rightarrow V$ defined by $O(v) = 0 \forall v \in V$ is the zero element of $\text{Hom}(V, V)$. Since for any $f \in \text{Hom}(V, V)$, $(f+O)(v) = f(v) + O(v) = f(v)$. This gives that $f+O = f$.

(4) Existence of additive inverse: For any $f \in \text{Hom}(V, V)$ define $-f: V \rightarrow V$ by $(-f)(v) = -f(v) \forall v \in V$.

$$\begin{aligned} \text{As for } u, v \in V, (-f)(u+v) &= -f(u+v) \\ &= -[f(u) + f(v)] = -f(u) + [-f(v)] \\ &= (-f)(u) + (-f)(v), -f \in \text{Hom}(V, V). \end{aligned}$$

$$\begin{aligned}\text{Further } [f+(-f)](v) &= f(v) + [-f(v)] \quad \forall v \in V \\ &= 0 \\ &= O(v).\end{aligned}$$

So $-f$ is the additive inverse of f .

(5) *Multiplication is Associative:* For all $f, g, h \in \text{Hom}(V, V)$

$$\begin{aligned}[(fg)h](v) &= (fg)[h(v)] \quad \forall v \in V \\ &= f\{g[h(v)]\}\end{aligned}$$

$$\text{Similarly } [f(gh)](v) = f\{g[h(v)]\}.$$

$$\text{Hence } (fg)h = f(gh).$$

(6) *Distributivity of multiplication over addition:*

For all $f, g, h \in \text{Hom}(V, V)$

$$\begin{aligned}[f(g+h)](v) &= f[(g+h)(v)] \quad \forall v \in V \\ &= f[g(v) + h(v)] \\ &= f[g(v)] + f[h(v)] \\ &= fg(v) + fh(v) \\ &= (fg + fh)(v) \\ &\Rightarrow f(g+h) = fg + fh.\end{aligned}$$

Similarly it can be proved that $(g+h)f = gf + hf$.

(7) *Existence of unity:* The identity mapping $I: V \rightarrow V$ is clearly a group homomorphism of V , so $I \in \text{Hom}(V, V)$. Further for any $f \in \text{Hom}(V, V)$, $(fI)(v) = f[I(v)] = f(v) \quad \forall v \in V$. So that we get $fI = f$.

Similarly we can show that $If = f$.

Hence I is unity of $\text{Hom}(V, V)$.

Consequently $\text{Hom}(V, V)$ is a ring with unity. ■

Theorem 8.27. Every ring with unity can be embedded in a ring of endomorphism of some additive Abelian group.

Proof: Let R be a ring with unity 1 and let V be the additive group R^+ of R .

Define $f: R \rightarrow \text{Hom}(V, V)$ by $[f(r)](v) = rv \quad \forall v \in V$. Clearly $rv \in V$ and $[f(r)](u+v) = r(u+v) = ru + rv = [f(r)](u) + [f(r)](v)$ for all $u, v \in V$. Thus $f(r) \in \text{Hom}(V, V)$ and so our definition is meaningful.

$$\begin{aligned}\text{Now for } r, s \in R, [f(r+s)](v) &= (r+s)v \quad \forall v \in V \\ &= rv + sv = [f(r)](v) + [f(s)](v) = [f(r) + f(s)](v) \\ &\Rightarrow f(r+s) = f(r) + f(s).\end{aligned}$$

$$\begin{aligned}\text{Further } [f(rs)](v) &= (rs)v = r(sv) \text{ as } r, s, v \in R \\ &= [f(r)](sv) = f(r)\{[f(s)](v)\} = [f(r)f(s)](v) \\ &\Rightarrow f(rs) = f(r)f(s). \text{ Hence } f \text{ is a homomorphism.}\end{aligned}$$

$$\begin{aligned}\text{Finally } f(r) = f(s) &\Rightarrow [f(r)](v) = [f(s)](v) \Rightarrow rv = sv \quad \forall v \in V \\ &\Rightarrow r1 = s1 \text{ as } 1 \in V = R^+, \text{ hence } f(r) = f(s) \Rightarrow r = s.\end{aligned}$$

Consequently f is 1-1 and so $R \cong f(R) \subseteq \text{Hom}(V, V)$. In other words R is embeddable in $\text{Hom}(V, V)$. ■

Corollary 8.28. Every ring can be embedded in a ring of endomorphism of some additive Abelian group.

Proof: Let R be a ring, then there exists a ring R_1 with unity and a monomorphism $f: R \rightarrow R_1$ (Theorem 8.19). Further there exists an Abelian additive group V and a monomorphism $g: R_1 \rightarrow \text{Hom}(V, V)$ (Theorem 8.27).

Thus $gf: R \rightarrow \text{Hom}(V, V)$ is a homomorphism. Since both f and g are 1-1, gf is also 1-1. This gives $R \cong gf(R) \subseteq \text{Hom}(V, V)$. Hence R is embeddable in $\text{Hom}(V, V)$. ■

REMARK. Henceforth we shall denote by R^+ , the additive group of ring R . Thus we have proved that a ring R with unity is embeddable in $\text{Hom}(R^+, R^+)$.

WORKED-OUT EXERCISES

Exercise 1. Show that $\text{Hom}(\mathbf{Z}^+, \mathbf{Z}^+) \cong \mathbf{Z}$.

Solution. Define $f: \mathbf{Z} \rightarrow \text{Hom}(\mathbf{Z}^+, \mathbf{Z}^+)$ by $f(n) = T_n$ for all $n \in \mathbf{Z}$, where $T_n(x) = nx \quad \forall x \in \mathbf{Z}^+$.

Clearly $T_n(x+y) = n(x+y) = nx + ny = T_n(x) + T_n(y)$
 $\Rightarrow T_n \in \text{Hom}(\mathbf{Z}^+, \mathbf{Z}^+)$. Thus our definition makes sense.

It can be easily checked that f is a ring homomorphism.

Suppose $f(n) = f(m)$, then $T_n = T_m \Rightarrow T_n(x) = T_m(x) \quad \forall x \in \mathbf{Z}^+$. In particular $T_n(1) = T_m(1) \Rightarrow n = m$. Hence f is 1-1.

Finally let $\theta \in \text{Hom}(\mathbf{Z}^+, \mathbf{Z}^+)$. Let $\theta(1) = k$. We show that $\theta = T_k = f(k)$.

If x is a positive integer, $\theta(x) = \theta(1+1+\dots x \text{ times})$
 $= \theta(1) + \theta(1) + \dots x \text{ times}$
 $= x\theta(1) = xk = kx = T_k(x)$.

If x is a negative integer $\theta(x) = \theta(-y)$, where y is a positive integer.

This gives that $\theta(x) = -\theta(y)$ as θ is a group homomorphism. So $\theta(x) = -ky = k(-y) = kx = T_k(x)$.

If $x=0$ then as θ is a group homomorphism $\theta(x) = 0 = k0 = T_k(0) = T_k(x)$.

Hence $\theta = T_k$. This shows that f is onto. Consequently $\mathbf{Z} \cong \text{Hom}(\mathbf{Z}^+, \mathbf{Z}^+)$. In other words $\text{Hom}(\mathbf{Z}^+, \mathbf{Z}^+) \cong \mathbf{Z}$.

Exercise 2. (Hua) A mapping ϕ from a ring R to a ring S is said to be an anti-homomorphism if for all $a, b \in R$ $\phi(a+b) = \phi(a) + \phi(b)$ and $\phi(ab) = \phi(b)\phi(a)$.

Let f be a mapping from a ring R to a ring S such that (i) $\forall x, y \in R, f(x+y) = f(x) + f(y)$ and (ii) for every pair x, y of R either $f(xy) = f(x)f(y)$ or $f(xy) = f(y)f(x)$.

Prove that f is either a homomorphism or an anti-homomorphism.

Solution. Let $W_c = \{x \in R \mid f(cx) = f(c)f(x)\}$

and $V_c = \{x \in R \mid f(cx) = f(x)f(c)\}.$

Clearly W_c and V_c are subgroups of $\langle R, + \rangle$ and for each $c \in R$ $R = W_c \cup V_c$ (By hypothesis). This will imply either $R = W_c$ or $R = V_c$.

Now suppose neither f is a homomorphism nor an anti-homomorphism. Since f is not a homomorphism, there exists $a \in R$ such that $R \neq W_a$. Hence we can find $b \in R - W_a$. This implies $f(ab) \neq f(a)f(b)$. Hence $f(ab) = f(b)f(a) \neq f(a)f(b)$.

But $R = W_a \cup V_a$ and $R \neq W_a \Rightarrow R = V_a$ so $\forall x \in R, f(ax) = f(x)f(a)$... (1)

We claim that $b \in V_y$ for all y in R . If not, $\exists z \in R$, such that $b \notin V_z$. This gives that $b \in W_z$. Hence $f(zb) = f(z)f(b) \neq f(b)f(z)$... (2)

Consider $(a+z)b$. If $f[(a+z)b] = f(a+z)f(b) = [f(a)+f(z)]f(b)$ we get that $f(a)f(b) + f(z)f(b) = f(ab+zb) = f(ab) + f(zb)$... (3)

$\Rightarrow f(ab) = f(a)f(b)$ by (3). This is against (1).

If $f[(a+z)b] = f(b)f(a+z) = f(b)[f(a)+f(z)]$ we get that $f(b)f(a) + f(b)f(z) = f(ab) + f(bz) \Rightarrow f(b)f(z) = f(bz)$ by (1). This is against (3).

Hence our claim is established. In other words $f(yb) = f(b)f(y) \forall y$ in R .

Since f is not an anti-homomorphism, there exists $c \in R$. Such that $R \neq V_c$. This implies the existence of d in R , satisfying $f(cd) = f(c)f(d) \neq f(d)f(c)$ (4)

Using same arguments as given earlier, we can prove that $f(cu) = f(c)f(u) \forall u \in R$... (5)

and $f(vd) = f(v)f(d) \forall v \in R$... (6)

Finally consider $f[(a+c)(b+d)]$ (7)

Case I. $f[(a+c)(b+d)] = f(b+d)f(a+c)$.

This implies that

$$f(ab+ad+cb+cd) = [f(b)+f(d)][f(a)+f(c)]$$

$$\Rightarrow f(ab)+f(ad)+f(cb)+f(cd) = f(b)f(a)+f(b)f(c)+f(d)f(a)+f(d)f(c).$$

By using (1), (2) and (4), we get $f(cd) = f(d)f(c)$, a contradiction to (5).

Case II. $f[(a+c)(b+d)] = f(a+c)f(b+d)$.

This gives that

$$f(ab+ad+cb+cd) = [f(a)+f(c)][f(b)+f(d)]$$

$$\Rightarrow f(ab)+f(ad)+f(cb)+f(cd) = f(a)f(b)+f(a)f(d)+f(c)f(b)+f(c)f(d).$$

Using (5), (6) and (7), we get $f(ab) = f(a)f(b)$ which is a contradiction to (1).

Hence our assertion that neither f is a homomorphism nor an anti-homomorphism is wrong. Hence the result follows.

Exercise 3. If R is a ring with unity 1 and f is a homomorphism of R into an integral domain R' . If $\text{Ker } f \neq R$, prove that $f(1)$ is unity

of R' .

Solution. First we claim that $f(1) \neq 0$, otherwise for all $x \in R$ $f(x) = f(1x) = f(1)f(x) = 0 \Rightarrow R = \text{Ker } f$, which is against hypothesis. Now $[f(1)]^2 = f(1)f(1) = f(1.1) = f(1)$.

Let $r \in R'$ then $[rf(1) - r]f(1) = r[f(1)]^2 - rf(1) = 0 \Rightarrow rf(1) = r$ as $f(1) \neq 0$ and R' is an integral domain. Since R' is commutative $f(1)r = r \forall r \in R'$. Hence $f(1)$ is unity of R' .

Exercise 4. Prove that the only homomorphisms of $\mathbb{Q}$ into itself are the identity mapping and the mapping that sends every element into zero.

Solution. Let $f: \mathbb{Q} \rightarrow \mathbb{Q}$ be a homomorphism. If $f(1) = 0$ then $\text{Ker } f = \mathbb{Q} \Rightarrow f(x) = 0 \forall x \in \mathbb{Q}$.

If $f(1) \neq 0$ then as $\mathbb{Q}$ is an integral domain, by Exercise 3, above, $f(1)$ is unity of $\mathbb{Q}$. So $f(1) = 1$.

Let n be a positive integer then $f(n) = f(1 + 1 + \dots n \text{ times}) = f(1) + f(1) + \dots n \text{ times} = n f(1) = n$.

If $n = 0$ then $f(0) = 0 \Rightarrow f(n) = n$. If n is a negative integer, put $n = -m$, where m is a positive integer. $f(-m) = -f(m) = -m$ by 1st case above. Hence $f(n) = n$.

Further for any rational number $\frac{p}{q}$, $p, q \in \mathbb{Z}$, $q \neq 0$

$$\begin{aligned} p &= q \frac{p}{q} \Rightarrow f(p) = f(q) f\left(\frac{p}{q}\right) \\ \Rightarrow f\left(\frac{p}{q}\right) &= \frac{f(p)}{f(q)} = \frac{p}{q}, \quad [\text{Note that } q \neq 0 \Rightarrow f(q) = q \neq 0]. \end{aligned}$$

Hence for all $x \in \mathbb{Q}$; $f(x) = x$, so $f = I$, the identity mapping on $\mathbb{Q}$.

Exercise 5. Let X be a set and f is a 1-1 mapping of X onto a ring R . Define $+$ and $\cdot$ on X as under.

For all $x, y \in X$, $x + y = f^{-1}[f(x) + f(y)]$, $x \cdot y = f^{-1}[f(x)f(y)]$.

Show that $\langle X, +, \cdot \rangle$ is a ring isomorphic to R .

Solution. Since $f(x), f(y) \in R$, $f^{-1}[f(x) + f(y)]$, $f^{-1}[f(x)f(y)] \in X$. As f is 1-1 onto mapping $f^{-1}[f(x) + f(y)]$, $f^{-1}[f(x)f(y)]$ are uniquely defined for any pair of elements $x, y \in X$.

Note that definitions of $+$ and $\cdot$ imply that $f(x + y) = f(x) + f(y)$ and $f(x \cdot y) = f(x)f(y) \forall x, y \in X$.

Now $\forall x, y, z \in X$.

$$\begin{aligned} f[(x + y) + z] &= f(x + y) + f(z) = f(x) + f(y) + f(z) \\ &= f(x) + f[(y + z)] \\ &= f[x + (y + z)] \end{aligned}$$

$$\Rightarrow (x + y) + z = x + (y + z).$$

Similarly it can be checked that $\forall x, y, z \in X$, $x + y = y + x$, $(x \cdot y) \cdot z = x \cdot (y \cdot z)$, $x \cdot (y + z) = x \cdot y + x \cdot z$ and $(y + z) \cdot x = y \cdot x + z \cdot x$.

$f^{-1}(0)$ works as zero for X , since $x + f^{-1}(0) = f^{-1}[f(x) + ff^{-1}(0)] = f^{-1}[f(x) + 0] = f^{-1}[f(x)] = x$. Also for each $x \in X$, $f^{-1}[-f(x)]$ works as negative. For

$$\begin{aligned} x + f^{-1}[-f(x)] &= f^{-1}\{f(x) + f f^{-1}[-f(x)]\} \\ &= f^{-1}\{f(x) - f(x)\} = f^{-1}(0). \end{aligned}$$

Consequently $\langle X, +, \cdot \rangle$ is a ring.

Clearly $f: X \rightarrow R$ is a homomorphism which is already given to be 1-1 and onto.

Hence $X \cong R$.

PROBLEMS

1. Show that $\text{Hom}(\mathbf{Z}_n^+, \mathbf{Z}_n^+) \cong \mathbf{Z}_n$ where $\mathbf{Z}_n$ is the ring of integers modulo n .
2. Show that a commutative ring R with unity is isomorphic to $\text{Hom}(R^+, R^+)$ if and only if $\text{Hom}(R^+, R^+)$ is commutative.
3. If $a, b \in D$ where D is an integral domain with at least two elements, such that $a^m = b^m$ and $a^n = b^n$ for two relatively prime integers m and n , prove that $a = b$.
4. Give an example to show that in the above problem if we replace integral domain by a commutative ring the result need not be true.

[Hint. In $\mathbf{Z}/(8)$, $4^3 = 2^3$ and $4^4 = 2^4$ but $4 \neq 2$.]

5. Let R be a commutative ring with unit. A non-void subset S of R is called a *multiplicative subset* if

(i) $0 \notin S$ and (ii) $s_1, s_2 \in S \Rightarrow s_1 s_2 \in S$.

Let $R \times S = \{(r, s) \mid r \in R, s \in S\}$. Define a relation ' $\sim$ ' on $R \times S$ as under:

$(r, s) \sim (r', s')$ if and only if there exists $s'' \in S$ such that $s''(rs' - sr') = 0$ for all $(r, s), (r', s') \in R \times S$. Prove that ' $\sim$ ' is an equivalence relation on $R \times S$ and the set $Q = \{r/s \mid r/s \text{ is equivalence class of } (r, s)\}$ is a ring under addition and multiplication defined by $r/s + r_1/s_1 = (rs_1 + r_1s)/ss_1$, $(r/s)(r_1/s_1) = rr_1/ss_1$ for all $r/s, r_1/s_1 \in Q$.

Define $f: R \rightarrow Q$ by $f(a) = as/s$ for some fixed $s \in S$. Show that f is a homomorphism of R into Q .

Lastly prove that every element of Q of the form s_1/s_2 , $s_1, s_2 \in S$ has an inverse in Q .

Q is called a *ring of quotients* of R with respect to S and we denote it by R_s .

6. *Definition.* An element of a ring R is said to be a *regular element* if it is neither a left zero-divisor nor a right zero-divisor. Let R be a commutative ring with unity and let U be the set of all

regular elements of R . Show that U is a multiplicative subset (see problem 5). The ring of quotients of R with respect to U is called *Classical Quotient Ring of R* (or *total ring of quotients of R*) and is denoted by $Q_{cl}(R)$. Prove the following:—

(i) If $a, b \in R$ then ab^{-1} is a regular element of $Q_{cl}(R)$ if and only if a is a regular element of R .

(ii) If every regular element of R is a unit of R then show that $R \cong Q_{cl}(R)$.

(iii) $Q_{cl}[Q_{cl}(R)] = Q_{cl}(R)$.

(iv) If R is finite then $R \cong Q_{cl}(R)$.

7. If $R = \{a + b\sqrt{3} \mid a, b \in \mathbb{Z}\}$. Show that R is an integral domain with unity. Obtain its field of quotients.

[Ans. Field of quotients of $R = \{x + \beta\sqrt{3}, x, \beta \in \mathbb{Q}\}$.]

8. (Ore) Let R be a ring (not necessarily commutative) with unity. Prove that R possesses a classical quotient ring if and only if it satisfies the following condition.

For all $a, b \in R$; b regular, there exist $c, d \in R$; d , regular, such that $ad = bc$.

9. Show that the set $S = \{n \in \mathbb{Z} \mid p \nmid n\}$, where p is a fixed prime number is a multiplicative set. Determine $\mathbb{Z}_S$, the ring of quotients of $\mathbb{Z}$ with respect to S .

10. Let P be a proper prime ideal of a commutative ring R . Take $S = R - P$, show that S is a multiplicative subset. Denote R_s by R_p . Show that PR_p is the unique maximal ideal of R_p , where $PR_p = \{(qs_1/s_1)(r/s) \mid q \in P, r \in R, s, s_1 \in S\}$.

11. Let S be a multiplicative subset of a commutative ring R with unity, prove that the mapping $f: R \rightarrow R_S$ given by $f(r) = rs/s$ for some fixed $s \in S$, is a monomorphism if and only if S contains no proper zero-divisors.

12. Let S be a multiplicative subset of a commutative ring R with unity. Prove

(i) $I = \{a \in R \mid as = 0 \text{ for some } s \in S\}$ is an ideal of R .

(ii) S/I is a multiplicative subset of R/I where $S/I = \{s + I \mid s \in S\}$.

(iii) No element of S/I is a zero-divisor of R/I .

[The ring of quotients of R/I with respect to S/I is called *generalized ring of quotients of R with respect to S* .]

(iv) $(R/I)_{S/I} \cong R_S$.

13. Let U be the set of all regular elements of a commutative ring R . Prove:

(a) If I is an ideal of R then $IU^{-1} = \{a/b \in Q_{cl}(R) \mid a \in I, b \in U\}$ is an ideal of $Q_{cl}(R)$. Conversely each ideal J of $Q_{cl}(R)$ is of the form $(J \cap R)U^{-1}$.

(b) For ideals A and B of R , $(A+B)U^{-1}=AU^{-1}+BU^{-1}$:
 $(AB)U^{-1}=(AU^{-1})(BU^{-1})$ and $(A\cap B)U^{-1}=AU^{-1}\cap BU^{-1}$.

14. Determine the field of quotients of the following integral domains.

(i) The domain of all rational numbers of the form $m/10^n$ (m, n integers).

(ii) The domain of all complex numbers $a+bi$ where $a, b \in \mathbb{Z}$.

[Ans. (i) $\mathbb{Q}$ (ii) $\{c+di \mid c, d \in \mathbb{Q}\}$.]

15. Give an example to show that non-isomorphic integral domains may have isomorphic fields of quotients.

[Hint. Consider $\mathbb{Z}$ and E , the ring of even integers]

POLYNOMIAL RINGS

1. Polynomials

In school every student who opts for Mathematics comes across the concept of polynomials, their addition, multiplication and factorization. Also he is taught to find the solutions of polynomial equations. Later, at the beginning of college education, the polynomial functions, their limits and continuity are also taught. Here we want to discuss the polynomials, but not essentially from any point of view as mentioned above, instead we shall treat polynomials as members of some ring and study the algebraic properties of that ring.

Consider any ring S and a subring R of S . Let $u \in S$ be such that $ru = ur$ for every $r \in R$. Let $R[u]$ denote the set of all elements of S of the form

$$a_0 + a_1u + a_2u^2 + \dots + a_nu^n,$$

where $a_i \in R$ for every $i=0, 1, 2, \dots, n$ and n is any positive integer. Using the fact that $au^k = u^ka$ for every $a \in R$ and every positive integer k , one can immediately see that $R[u]$ is non empty and closed under subtraction and multiplication. That is, $R[u]$ is a subring of S . Further if

$$a_0 + a_1u + a_2u^2 + \dots + a_nu^n = 0$$

for $a_i \in R$ ($i=0, 1, 2, \dots, n$) such that at least one of a_i is non-zero then we say that u is algebraic over R . However if for every choice of $a_0, a_1, a_2, \dots, a_n$ (of elements in R) $a_0 + a_1u + a_2u^2 + \dots + a_nu^n \neq 0$ whenever at least one of $a_i \neq 0$ ($i > 0$) then u is said to be transcendental over R . For example let us consider $\mathbf{R}$ and $\mathbf{Q}$. Now $\sqrt{2} \in \mathbf{R}$, by using the fact that $(\sqrt{2})^2 = 2 \in \mathbf{Q}$, we can see that each member of $\mathbf{Q}[\sqrt{2}]$ is of the type $a + b\sqrt{2}$, $a, b \in \mathbf{Q}$. Further as $2 - 1(\sqrt{2})^2 = 0$, we see that $\sqrt{2}$ is algebraic over $\mathbf{Q}$. Consider π , we know that π is not an algebraic number i.e. it is not the root of any polynomial equation with coefficients as rational numbers. Thus π is transcendental over $\mathbf{Q}$.

Given any member

$$a_0 + a_1\pi + a_2\pi^2 + \dots + a_n\pi^n \in \mathbf{Q}[\pi]$$

we get an infinite sequence $(a_0, a_1, a_2, \dots, a_n, 0, 0, \dots)$ of elements of $\mathbf{Q}$ in which all terms after n th stage are equal to zero. Conversely given any infinite sequence

$$(b_0, b_1, \dots, b_m, 0, 0, \dots)$$

of elements of $\mathbf{Q}$ in which all except finite number of terms are zero we can find a non-negative integer m such that $b_i = 0 \forall i \geq m+1$. This sequence determines an element

$$b_0 + b_1\pi + b_2\pi^2 + \dots + b_m\pi^m \in \mathbf{Q}[\pi].$$

Further notice that any two elements

$$c_0 + c_1\pi + c_2\pi^2 + \dots \text{ and } d_0 + d_1\pi + d_2\pi^2 + \dots$$

with $c_i, d_i \in \mathbf{Q}$, are equal if and only if $c_i = d_i \forall i$, since otherwise we shall get π to be algebraic over $\mathbf{Q}$. Thus each member $a_0 + a_1\pi + a_2\pi^2 + \dots + a_n\pi^n$ is uniquely determined by the corresponding infinite sequence $(a_0, a_1, a_2, \dots, a_n, 0, 0, \dots)$ of elements of $\mathbf{Q}$. Again we turn to the general case of R and S . In that case if $u \in S$ is transcendental over R then each member $a_0 + a_1u + \dots + a_nu^n, a_i \in R$ is uniquely determined by the infinite sequence $(a_0, a_1, a_2, \dots, a_n, 0, 0, \dots)$ of elements of R .

Let us take a look on the problem from different angle. Suppose R is a ring. For the sake of convenience let us suppose that R has unity $1 (\neq 0)$. The problem is 'does there exist a ring S containing R such that S contains an element x which commutes with every element of R and x is transcendental over R ?' It is this problem we take up firstly in this section and give its answer in affirmative. We take clue from our earlier observation that if x is transcendental over R then any member $a_0 + a_1x + a_2x^2 + \dots + a_nx^n$ determines uniquely and is determined uniquely by the infinite sequence $(a_0, a_1, a_2, \dots, a_n, 0, 0, \dots)$ of elements of R . As we do not know for the time being, what this x will be—we prefer to define a polynomial over R as follows:

Definition 9.1. Any infinite sequence $(a_0, a_1, a_2, \dots, a_n, a_{n+1}, \dots)$ of elements of a ring R is said to be a *polynomial over R* if all except finite number of its terms a_i are equal to zero.

In other words an infinite sequence $(a_0, a_1, a_2, \dots, a_n, \dots)$ of elements of R is said to be a polynomial over R if there exists a non-negative integer n such that $a_i = 0 \forall i \geq n+1$.

Each member a_j of the polynomial $(a_0, a_1, a_2, \dots)$ is called its *j th coefficient*. Further a_0 is called the *constant term* and if n is the largest non-negative integer with $a_n \neq 0$ then a_n is called the *leading coefficient* of the polynomial.

Definition 9.2. (Sum and product of two polynomials).

Let $f=(a_0, a_1, a_2, \dots)$ and $g=(b_0, b_1, b_2, \dots)$ be any two polynomials over a ring R then their sum $f+g$ and product fg are defined as under:

$$\begin{aligned} f+g &= (a_0+b_0, a_1+b_1, a_2+b_2, \dots) \\ fg &= (c_0, c_1, c_2, \dots) \text{ where } c_0=a_0+b_0, \\ c_1 &= a_0b_1+a_1b_0, c_2=a_0b_2+a_1b_1+a_2b_0 \text{ and so on, in general} \\ c_k &= a_0b_k+a_1b_{k-1}+\dots+\dots+a_kb_0 \\ &= \sum_{i+j=k} a_ib_j \text{ for every } k \geq 0. \end{aligned}$$

It remains to settle whether sum and product of two polynomials over R are polynomials over R or not. This is answered by the following more comprehensive theorem.

Theorem 9.3. The set T of all polynomials over a ring R is a ring under the addition and multiplication operation defined as follows:

For all $f=(a_0, a_1, a_2, \dots)$, $g=(b_0, b_1, b_2, \dots) \in T$

$$f+g=(a_0+b_0, a_1+b_1, a_2+b_2, \dots)$$

$$\text{and } fg=(c_0, c_1, c_2, \dots)$$

$$\text{where } c_k = \sum_{i+j=k} a_ib_j \text{ for every } k \geq 0.$$

Proof: Let a_n and b_m be coefficients of f and g respectively such that $a_i=0 \forall i \geq n+1$ and $b_j=0 \forall j \geq m+1$. If $s=\max(m, n)$ then $a_t=b_t=0 \forall t > s$, so $a_t+b_t=0 \forall t > s$, hence $f+g \in T$.

Again consider $k > m+n$, then if $i+j=k$ either $i > m$ or $j > n$ hence $a_ib_j=0$ for all i and j satisfying $i+j=k$ for $k > m+n$. This gives that $c_k=0$ for all $k \geq m+n+1$, hence $fg \in T$.

Now as R is a ring it can be easily verified that $\langle T, + \rangle$ is an Abelian group in which $(0, 0, 0, \dots)$ is the additive identity and for any $f=(a_0, a_1, a_2, \dots) \in T$, $f'=(-a_0, -a_1, -a_2, \dots)$ is the additive inverse of f . Clearly $f' \in T$. Further $\langle T, \cdot \rangle$ is a semi-group, as $\langle R, \cdot \rangle$ is a semi-group. Finally it can be easily checked that $\cdot$ is distributive on left as well as on right over $+$. Hence $\langle T, +, \cdot \rangle$ is a ring. ■

Theorem 9.4. Let T be a ring of polynomials over a ring R then $R'=\{(a, 0, 0, \dots) \mid a \in R\}$ is a subring of T isomorphic to R under the mapping $(a, 0, 0, \dots) \rightarrow a$. Further if R has unity 1 then $(1, 0, 0, \dots)$ is the unity of T .

Proof: Let $f: R \rightarrow T$ be defined by $f(a)=(a, 0, 0, \dots) \forall a \in R$. Then for all $a, b \in R$, $f(a+b)=(a+b, 0, 0, \dots)=(a, 0, 0, \dots) + (b, 0, 0, \dots)=f(a)+f(b)$ and $f(ab)=(ab, 0, 0, \dots)=(a, 0, 0, \dots)(b, 0, 0, \dots)=f(a)f(b)$. Also f is 1-1 since $f(a)=f(b) \Rightarrow (a, 0, 0, \dots)=(b, 0, 0, \dots) \Rightarrow a=b$.

Hence $R \cong f(R)=R'$. This proves first part of the theorem.

Finally if R has unity 1 then $f(1) = (1, 0, 0, \dots) \in T$ and for all $(a_1, a_2, a_3, \dots) \in T$, $(a_1, a_2, a_3, \dots)(1, 0, 0, \dots) = (a_1, a_2, a_3, \dots) = (1, 0, 0, \dots)(a_1, a_2, a_3, \dots)$. Hence $(1, 0, 0, \dots)$ is the unity of T . ■

REMARK. For each $a \in R$, let $\bar{a}$ denote the polynomial $(a, 0, 0, \dots)$. Further let R have unity $1 (\neq 0)$.

Define $x = (0, 1, 0, 0, \dots)$ then $\bar{a}x = (a, 0, 0, \dots)(0, 1, 0, 0, \dots) = (0, a, 0, 0, \dots) = (0, 1, 0, 0, \dots)(a, 0, 0, \dots) = x\bar{a}$ and so x commutes with every element of R' i.e. with every element of R if we identify each a with $\bar{a}$.

By applying the formula for multiplication of the polynomials one finds that $x^2 = (0, 0, 1, 0, 0, \dots)$, $x_3 = (0, 0, 0, 1, 0, 0, \dots)$ (n+1)th place

and so on, in general $x^n = (0, 0, 0, \dots, 0, 1, 0, \dots)$ (n+1)th place. Further

for any $a \in R$, $\bar{a}x^n = (0, 0, 0, \dots, a, 0, \dots)$.

Let $f = (a_0, a_1, a_2, \dots)$ be any polynomial over R . As there exists a non-negative integer n such that $a_i = 0 \forall i \geq n+1$ we see that $(a_0, a_1, a_2, \dots, a_n, 0, 0, \dots) = (a_0, 0, 0, \dots) + (0, a_1, 0, 0, \dots) + (0, 0, a_2, 0, \dots) + \dots + (0, 0, \dots, a_n, 0, \dots)$ (n+1)th place

$\dots = \bar{a}_0 + \bar{a}_1x + \bar{a}_2x^2 + \dots + \bar{a}_nx^n$.

Further $f=0$ if and only if $a_i=0 \forall i$ gives $\bar{a}_0 + \bar{a}_1x + \bar{a}_2x^2 + \dots + \bar{a}_nx^n = 0$ if and only if each $a_i=0$. As $R \cong R'$ under the mapping $a \rightarrow \bar{a}$ we identify R with R' and each $a \in R$ with the polynomial $\bar{a}$. Then we get that R is a subring of T . Now $x \in T$ is such that $ax = xa \forall a \in R$. Every member of T is of the form

$$f = a_0 + a_1x + a_2x^2 + \dots + a_nx^n$$

and $f=0$ only when each $a_i=0$. Consequently x is transcendental over R and $T = R[x]$. Thus we have found a ring T containing R such that T is a ring of polynomials in an element x over R and x is transcendental over R . We show that such ring T is unique to within isomorphism. For this we have.

Theorem 9.5. Let R be any ring and, S and S' two overrings of R . Suppose that there exist $u \in S, u' \in S'$ such that u and u' commute with every element of R . Then $R[u] \cong R[u']$ under the mapping $f: R[u] \rightarrow R[u']$ defined by $f(a_0 + a_1u + a_2u^2 + \dots) = a_0 + a_1u' + a_2(u')^2 + \dots$ whenever u and u' are transcendental over R .

Proof: Firstly we show that f is well-defined.

Let $a_0 + a_1u + a_2u^2 + \dots = b_0 + b_1u + b_2u^2 + \dots$

then $(a_0 - b_0) + (a_1 - b_1)u + (a_2 - b_2)u^2 + \dots = 0$

$\Rightarrow a_0 = b_0, a_1 = b_1, a_2 = b_2, \dots$ as u is transcendental over R .

This in turn implies $a_0 + a_1u' + a_2(u')^2 + \dots = b_0 + b_1u' + b_2(u')^2 + \dots$
hence $f(a_0 + a_1u + a_2u^2 + \dots) = f(b_0 + b_1u + b_2u^2 + \dots)$.

Consequently f is well-defined.

Now let $a_0 + a_1u + a_2u^2 + \dots, b_0 + b_1u + b_2u^2 + \dots \in R[u]$

Then $f[(a_0 + a_1u + a_2u^2 + \dots) + (b_0 + b_1u + b_2u^2 + \dots)]$
 $= f[(a_0 + b_0) + (a_1 + b_1)u + (a_2 + b_2)u^2 + \dots]$
 $= (a_0 + b_0) + (a_1 + b_1)u' + (a_2 + b_2)(u')^2 + \dots$
 $= [a_0 + a_1u' + a_2(u')^2 + \dots] + [b_0 + b_1u' + b_2(u')^2 + \dots]$
 $= f(a_0 + a_1u + a_2u^2 + \dots) + f(b_0 + b_1u + b_2u^2 + \dots)$.

Further $f[(a_0 + a_1u + a_2u^2 + \dots)(b_0 + b_1u + b_2u^2 + \dots)]$
 $= f[a_0b_0 + (a_0b_1 + a_1b_0)u + (a_0b_2 + a_1b_1 + a_2b_0)u^2 + \dots]$
 $= a_0b_0 + (a_0b_1 + a_1b_0)u' + (a_0b_2 + a_1b_1 + a_2b_0)(u')^2 + \dots$
 $= [a_0 + a_1u' + a_2(u')^2 + \dots][b_0 + b_1u' + b_2(u')^2 + \dots]$
 $= f(a_0 + a_1u + a_2u^2 + \dots)f(b_0 + b_1u + b_2u^2 + \dots)$.

Thus f is a homomorphism.

Again $f(a_0 + a_1u + a_2u^2 + \dots) = f(b_0 + b_1u + b_2u^2 + \dots)$

$\Rightarrow a_0 + a_1u' + a_2(u')^2 + \dots = b_0 + b_1u' + b_2(u')^2 + \dots$

$\Rightarrow (a_0 - b_0) + (a_1 - b_1)u' + (a_2 - b_2)(u')^2 + \dots = 0$

$\Rightarrow a_0 = b_0, a_1 = b_1, a_2 = b_2, \dots$, as u is transcendental over R .

$\Rightarrow a_0 + a_1u + a_2u^2 + \dots = b_0 + b_1u + b_2u^2 + \dots$

This gives f is 1-1.

Finally each element of $R[u']$ is of the type $a_0 + a_1u' + a_2(u')^2 + \dots$ with $a_0, a_1, a_2, \dots \in R$, which is clearly equal to $f(a_0 + a_1u + a_2u^2 + \dots)$.

Hence f is onto.

Consequently $R[u] \cong R[u']$. ■

Let us return to R and T mentioned earlier. We call $x = (0, 1, 0, 0, \dots, \dots)$ an *indeterminate* over R . Now if R has no unity then R can be embedded in a ring S with unity $1 (\neq 0)$. Then $T \subseteq S[x]$ with $x = (0, 1, 0, 0, \dots, \dots) \in S[x]$ but $x \notin T$. However each member of T is still of the form $a_0 + a_1x + a_2x^2 + \dots + a_nx^n, a_i \in R$.

Still we denote T by $R[x]$. In this case it must be emphasized that $x \notin R[x]$.

Theorem 9.6. (i) If R is commutative then $R[x]$ is commutative.

(ii) If R has no proper zero divisors then $R[x]$ also has no proper zero-divisors.

Proof: (i) Let $f = (a_0, a_1, a_2, \dots, \dots)$, $g = (b_0, b_1, b_2, \dots, \dots) \in R[x]$, then $fg = (c_0, c_1, c_2, \dots, \dots)$ where $c_k = a_0b_k + a_1b_{k-1} + \dots + \dots + a_kb_0 \forall k \geq 0$.

Also $gf = (d_0, d_1, d_2, \dots, \dots)$

where $d_k = b_0 a_k + b_1 a_{k-1} + \dots + b_k a_0 \forall k \geq 0$.

As R is commutative $a_i b_{k-i} = b_{k-i} a_i \forall i = 0, 1, \dots, k$.

Hence $c_k = d_k \forall k \geq 0$.

i.e. $fg = gf$. Consequently $R[x]$ is commutative.

(ii) Let $f \neq 0, g \neq 0$ then there exist non-negative integers m, n such that $a_n \neq 0, a_i = 0 \forall i \geq n+1$ and $b_m \neq 0, b_j = 0 \forall j \geq m+1$.

Let $fg = (c_0, c_1, c_2, \dots, \dots)$.

Then $c_{m+n} = a_0 b_{m+n} + a_1 b_{m+n-1} + \dots + a_{n-1} b_{m+1} + a_n b_m + a_{n+1} b_{m-1} + \dots + a_{m+n} b_0$
 $= a_n b_m$ since $a_i = 0 \forall i \geq n+1$ & $b_j = 0 \forall j \geq m+1$
 $\neq 0$ as $a_n \neq 0, b_m \neq 0$ and R has no proper zero-divisors.

Consequently $fg \neq 0$. This proves that R has no proper zero-divisors. ■

Corollary 9.7. If R is an integral domain then $R[x]$ is also an integral domain. ■

Definition 9.8. Let $f = (a_0, a_1, a_2, \dots, \dots)$ be any non-zero polynomial over a ring R . Then *degree* of f , denoted by $\deg f$ is equal to the largest non-negative integer n such that a_n , the n th coefficient of f is non-zero.

Thus if f is a non-zero polynomial, we know that f has at least one non-zero coefficient, further as f has only finite number of non-zero coefficients, we can find a non-negative integer n such that $a_n \neq 0$ but $a_i = 0 \forall i \geq n+1$.

Then $f = (a_0, a_1, a_2, \dots, \dots, a_n, 0, 0, \dots, \dots)$
 $= a_0 + a_1 x + a_2 x^2 + \dots + a_n x^n$, and $\deg f = n$.

We emphasize that we have not defined degree of zero polynomial. Of course some authors define the degree of zero polynomial to be equal to $-\infty$.

Theorem 9.9. For any two non-zero polynomials $f = (a_0, a_1, a_2, \dots, \dots)$, $g = (b_0, b_1, b_2, \dots, \dots)$ over a ring R ,

(i) if $f+g \neq 0$, then $\deg(f+g) \leq \max(\deg f, \deg g)$.

(ii) if $fg \neq 0$ then $\deg(fg) \leq \deg f + \deg g$.

(iii) if R is a ring without proper zero-divisors then $\deg fg = \deg f + \deg g$.

Proof: Let $\deg f = n$ and $\deg g = m$, then $a_n \neq 0, b_m \neq 0, a_i = 0 \forall i > n, b_j = 0 \forall j > m$.

(i) By definition of addition $f+g = (c_0, c_1, c_2, \dots, \dots)$
 where $c_i = a_i + b_i \forall i \geq 0$.

Now for $i > \max(n, m)$, we have $i > m, i > n$
 so $a_i = 0, b_i = 0$ and hence $c_i = 0$.

Then by definition $\deg(f+g) \leq \max(n, m)$.

(ii) Let $fg = (d_0, d_1, d_2, \dots, \dots) \neq 0$

with $d_k = \sum_{i+j=k} a_i b_j \forall k \geq 0$.

If $k > m+n$ then for any two non-negative integers i, j ; $i+j=k \Rightarrow i > n$ or $j > m$.

This gives $a_i=0$ or $b_j=0$ and so in each case $a_i b_j=0$. Hence $d_k=0 \forall k > m+n$.

Thus $\deg(fg) \leq m+n = \deg f + \deg g$.

(iii) We know that $d_{n+m} = a_n b_m \neq 0$ when R is a ring without proper zero-divisors. Consequently $\deg(fg) \geq m+n$. As seen in (ii), $\deg(fg) \leq m+n$, we conclude that $\deg(fg) = \deg f + \deg g$.

We now give examples to show that in certain cases equality does not hold for parts (i) and (ii) of the above theorem.

EXAMPLE 1. Consider the ring $R = \mathbb{Z}/(6)$. Its elements are $\bar{0}, \bar{1}, \bar{2}, \bar{3}, \bar{4},$ and $\bar{5}$. Take $f = \bar{1} + \bar{2}x$ and $g = \bar{5} + \bar{4}x + \bar{3}x^2$. These are polynomials in $R[x]$ of degree 1 and 2 respectively.

Now $fg = \bar{5} + \bar{2}x + \bar{5}x^2$ is of degree 2. In this case $\deg fg = 2$ which is not equal to $\deg f + \deg g$.

EXAMPLE 2. Consider the polynomials $f = 1 + 2x - 3x^2$ and $g = 4 - 3x + 3x^2$ in $\mathbb{Z}[x]$. Now $\deg f = 2$ and $\deg g = 2$. But $f+g = 5-x$ is of degree 1. In this case $\deg(f+g)$ is not equal to $\max(\deg f, \deg g)$.

REMARK. For any ring R if $f = a_0 + a_1x + a_2x^2 + \dots + a_nx^n$ and $g = b_0 + b_1x + b_2x^2 + \dots + b_mx^m$ are two polynomials of degree n and m respectively such that a_n or b_m is not a zero-divisor then $\deg fg = \deg f + \deg g$. This is so, because in fg the coefficient of x^{n+m} is $a_nb_m \neq 0$ as one of a_n and b_m is not a zero-divisor.

2. Factorization in $R[x]$

It turns out that the theory of factorization of integers in terms of prime numbers has a strong analogue in the theory of factorization of polynomials over a field. This analogue will be studied in the next chapter within a more general setting. But we end this chapter with the theorem of division algorithm in $R[x]$, where R is a commutative ring with unity. We shall see that this theorem plays a leading role in the development of factorization theory in $F[x]$ where F is a field, and also in the general theory of fields.

Theorem 9.10. (Division Algorithm in $R[x]$).

Let R be a commutative ring with unity; g , a non-zero polynomial in $R[x]$ of degree n with leading coefficient, a unit in R , then for any $f \in R[x]$ there exist unique polynomials h and r in $R[x]$ such that $f = hg + r$ where either $r = 0$ or $\deg r < \deg g$.

Proof: Firstly we establish the existence of h and r . If $f = 0$

then we can take $h=0, r=0$ and the result holds trivially.

Now let f be non-zero. We apply induction on $\deg f$ and use the fact that result holds for $f=0$, as seen above. Let the result hold for all polynomials of degree less than that of f and let $\deg f=m$.

In case $m < n$, taking $h=0$ and $r=f$ we get $f=hg+r$. Suppose $m \geq n$, then we can write

$$f = b_0 + b_1x + b_2x^2 + \dots + b_mx^m; b_m \neq 0 \quad \dots (1)$$

$$g = a_0 + a_1x + a_2x^2 + \dots + a_nx^n; a_n \text{ is a unit in } R. \dots (2)$$

Since a_n^{-1} exists and $m \geq n$, (2) yields

$$b_ma_n^{-1}x^{m-n}g = b_ma_n^{-1}a_0x^{m-n} + b_ma_n^{-1}a_1x^{m-n+1} + \dots + b_ma_n^{-1}a_{n-1}x^{m-1} + b_mx^m. \quad \dots (3)$$

By subtracting (3) from (1) and by using the fact that both polynomials are of same degree m and have same leading term b_m , we get that $f - b_ma_n^{-1}x^{m-n}g$ is a polynomial which is either zero or is of degree $< m$. In case the polynomial is zero, $h = b_ma_n^{-1}x^{m-n}, r=0$ satisfy the given conditions. In case this polynomial is not zero, by the induction hypothesis there exist $h_1, r \in R[x]$ such that

$$f - b_ma_n^{-1}x^{m-n}g = h_1g + r \quad \dots (4)$$

where either $r=0$ or $\deg r < \deg g$.

(4) immediately gives that $f = hg + r$.

with $h = b_ma_n^{-1}x^{m-n} + h_1$.

This establishes the first part of the theorem. \dots (5)

Now we show the uniqueness of h and r .

Let $f = hg + r = h_1g + r_1$ with $h, h_1, r, r_1 \in R[x]$ satisfying the given conditions. Then $(h-h_1)g = r_1-r$. \dots (6)

If $h-h_1 \neq 0$ then $\deg [(h-h_1)g] = \deg (h-h_1) + \deg g \geq \deg g$ and also $r_1-r \neq 0$ (by the remark following Example 2). But then $\deg (r_1-r) < \deg g$, which is a contradiction to equation (6). Hence $h=h_1$ and consequently $r=r_1$. ■

Since in a field every non-zero element is a unit, the above theorem gives.

Corollary 9.11. *If F is a field, then for any two polynomials $f, g \in F[x], g \neq 0$, there exist unique polynomials $h, r \in F[x]$ such that $f = hg + r$ and either $r=0$ or $\deg r < \deg g$. ■*

REMARK. The polynomials h and r of the above theorem are called the *quotient* and the *remainder* respectively while f and g are called *dividend* and *divisors* respectively. In case $r=0$ we say that g divides f or g is a factor of f or f is divisible by g . Symbolically $g \mid f$.

PROBLEMS

1. Complete the proof of Theorem 9.3.

[Hint. The only difficult steps in the proof are the verification of associativity of multiplication and distributivity of multiplication over addition. Let $f=(a_0, a_1, a_2, \dots, \dots, \dots)$, $g=(b_0, b_1, b_2, \dots, \dots)$ and $h=(c_0, c_1, c_2, \dots, \dots) \in T$.

To prove $(fg)h=f(gh)$ we proceed as under:

The k th component of fg is $d_k = \sum_{i+j=k} a_i b_j$.

Further r th component of $(fg)h$ is $\sum_{l+k=r} d_k c_l$

$$= \sum_{l+k=r} \left(\sum_{i+j=k} a_i b_j \right) c_l = \sum_{i+j+l=r} (a_i b_j) c_l,$$

where the last sum runs over all triples of integers $0 \leq i, j, l \leq r$ satisfying the equation $i+j+l=r$.

Similarly the r th component of $f(gh)$ is equal to

$$\sum_{i+k=r} a_i \left(\sum_{j+l=k} b_j c_l \right) = \sum_{i+j+l=r} a_i (b_j c_l).$$

This gives $(fg)h=f(gh)$.

Now to prove $f(g+h)=fg+fh$, note that the r th component of $f(g+h)$ is $\sum_{i+j=r} a_i (b_j + c_j)$ whereas the r th component of $fg+fh$ is

$$\sum_{i+j=r} a_i b_j + \sum_{i+j=r} a_i c_j. \text{ Hence } f(g+h)=fg+fh.]$$

2. Let R be a ring. Verify that the set S of all those polynomials in $R[x]$ which have constant term zero, form a subring of $R[x]$.

3. Let R be a ring with unity and $A = \langle x^2 \rangle$ be the ideal of $R[x]$ generated by x^2 . Show that $\frac{R[x]}{A} = \left\{ \frac{a_0 + a_1 x}{A} \mid a_0, a_1 \in R \right\}$. Hence

conclude that if R is a finite ring then $\frac{R[x]}{A}$ is also a finite ring.

4. For any ideal A of a ring R show that $A[x]$ is an ideal of $R[x]$. Prove that $\frac{R[x]}{A[x]} \cong \frac{R}{A} [x]$.

5. Let A be an ideal of a ring R . Prove that the set of all those polynomials in $R[x]$, whose constant terms are arbitrary but all other coefficients belong to A , is a subring of $R[x]$.

6. If S and T are subrings of R such that $ST \subseteq T$ and $TS \subseteq T$, prove that the set of all those polynomials in $R[x]$, whose constant terms are in S and other coefficients are in T , is a subring of $R[x]$.

7. Let R be a commutative ring with unity. Show that every polynomial $f \in R[x]$ is a unit if and only if the constant term of f is a unit and all other coefficients are nilpotent.

8. Let R be a commutative ring with unity. For all polynomial $f = a_0 + a_1 x + a_2 x^2 + \dots + a_n x^n \in R[x]$; A_f , the ideal of R generated by the elements $a_0, a_1, a_2, \dots, a_n$ is called the content of f . Using the

fact that every proper ideal of R is contained in a maximal ideal, prove the following:

If for two polynomials f and g , $A_f = R$, $A_g = R$ then $A_{fg} = R$.

9. Let f and g be two non-zero polynomials over a ring R . Prove the following:

(a) If $\deg f \neq \deg g$ then $f+g \neq 0$ and $\deg(f+g) = \max(\deg f, \deg g)$.

(b) If $\deg f = \deg g$ and $f+g \neq 0$ then $\deg(f+g) \leq \deg f$.

10. Prove the following generalized version of the division algorithm.

Let R be any commutative ring with unity. Let $f, g \in R[x]$ with $g \neq 0$ and α be the leading coefficient of g , prove that there exist polynomials h and r in $R[x]$ and a positive integer k such that

$$\alpha^k f = hg + r \text{ with } r = 0 \text{ or } \deg r < \deg g.$$

If α is not a zero-divisor prove that for a fixed k satisfying the given conditions, h and r are unique. How does this result generalize the division algorithm?

11. Let R be a commutative ring. If $f = a_0 + a_1x + a_2x^2 + \dots + a_nx^n$ in $R[x]$ is a zero-divisor, prove that there is an element $b(\neq 0)$ in R such that $ba_0 = ba_1 = \dots = ba_n = 0$.

[Hint. Choose a polynomial $g = c_0 + c_1x + \dots + c_mx^m$ of least degree m such that $fg = 0$. Then $a_n c_m = 0$. Consider $a_n g$, if it is not zero, its degree is less than m and still $f(a_n g) = 0$. This shows that $a_n g = 0$. Apply induction on r to prove that $a_{n-r}g = 0$ for all $0 \leq r \leq n$.]

12. Let R be a commutative ring with unity and let S be a subring of R that contains the unity of R and let $f \in R[x]$. If there exists a polynomial $g \in S[x]$ whose leading coefficient is a unit of S such that $fg \in S[x]$ then prove that $f \in S[x]$.

13. For any ring R show that $\frac{R[x]}{\langle x \rangle} \cong R$.

[Hint. Define $f: R[x] \rightarrow R$ by $f(a_0 + a_1x + a_2x^2 + \dots + a_nx^n) = a_0$.]

14. Show that the additive group of $Z[x]$ is isomorphic to the group G of positive rational numbers under multiplication.

[Hint. Use the fact that the set of prime numbers is countable.

Define $f: Z[x] \rightarrow G$ by $f\left(\sum_{i=0}^n a_i x^i\right) = \prod_{i=0}^n p_i^{a_i}$.

10

FACTORIZATION THEORY IN INTEGRAL DOMAINS

In Number theory one comes across the Fundamental Theorem of Arithmetic which states that any integer $n > 1$ is expressible as a product of powers of distinct primes and this expression is unique to within order. In this chapter we intend to generalize this result to the elements of certain special domains whose definitions are motivated by the properties of ring of integers. In this chapter we shall be mainly concerned with commutative rings with unity.

1. Divisibility

Definition 10.1. Let R be a commutative ring and let a and b be two elements of R , b is said to *divide* a (symbolically $b \mid a$) if $a = bc$ for some $c \in R$. b is called a *factor* of a . Further b is said to be a *proper factor* of a if b and c both are non-units.

The proof of the following theorem is trivial and hence omitted.

Theorem 10.2. Let R be a commutative ring with unity and let a, b, c be any three elements of R , then (i) $a \mid a$.

(ii) If $a \mid b$ and $b \mid c$ then $a \mid c$ and $a \mid cx$ for every $x \in R$.

(iii) If $a \mid b$ and $a \mid c$ then $a \mid (b \pm c)$. ■

REMARK. By definition it follows that every element of a ring R is a factor of 0.

Definition 10.3. Let R be a commutative ring with unity. An element a of R is said to be an *associate* of $b \in R$ if $a = bu$ for some unit $u \in R$.

REMARK. $a \sim b$ will denote that a is an associate of b .

Theorem 10.4. Let R be a commutative ring with unity, then the following hold:

(i) The relation of being associates is an equivalence relation on R .

(ii) If R is a domain and a, b are two non-zero elements of R then

$a \sim b$ if and only if $a \mid b$ and $b \mid a$.

(iii) If a, b are two non-zero elements of R then $a \mid b$ and $b \mid a$ if and only if $\langle a \rangle = \langle b \rangle$.

Proof: (i) For all $x \in R$, $x = 1x$ i.e. $x \sim x$, so $\sim$ is reflexive.

If $a \sim b$ then $a = bu$ for some unit $u \in R \Rightarrow b = au^{-1} \Rightarrow b \sim a$ as u^{-1} is again a unit in R .

Thus $\sim$ is symmetric.

Finally if $a \sim b$ and $b \sim c$ then $a = tu$ and $b = cv$ for some units $u, v \in R$. This gives $a = cvu$; however u, v being units in R implies that vu is again a unit in R . Thus $a \sim c$ and this implies that $\sim$ is transitive.

Hence $\sim$ is an equivalence relation on R .

(ii) If $a \sim b$ then $a = bu$ for some unit $u \in R \Rightarrow b \mid a$. Also $b = au^{-1} \Rightarrow a \mid b$.

Conversely if $a \mid b$ and $b \mid a$ we get $b = ac$ and $a = bd$ for some $c, d \in R$; thus $b = bdc \Rightarrow dc = 1$ as $b \neq 0$ and R is an integral domain; thus d is a unit. Hence $a \sim b$.

(iii) $a \mid b \Rightarrow b = ac$ for some $c \in R \Rightarrow b \in \langle a \rangle \Rightarrow \langle b \rangle \subseteq \langle a \rangle$. Similarly $b \mid a \Rightarrow \langle a \rangle \subseteq \langle b \rangle$, hence $\langle b \rangle = \langle a \rangle$.

Conversely $\langle a \rangle = \langle b \rangle \Rightarrow a \in \langle b \rangle \Rightarrow a = br$ for some $r \in R \Rightarrow b \mid a$. In the same fashion $\langle b \rangle = \langle a \rangle$ gives $a \mid b$. ■

Note. The only associate of 0 is zero.

EXAMPLE 1. In $\mathbb{Z}$, -3 and 3 are associates as -1 is a unit. In fact the only units in $\mathbb{Z}$ are $1, -1$; so every non-zero integer n has two associates namely $-n$ and n .

EXAMPLE 2. In $\mathbb{Z}/(8)$ only units are $\bar{1}, \bar{3}, \bar{5}$ and $\bar{7}$. The associates of $\bar{2}$ are therefore $\bar{2}$ and $\bar{6}$.

EXAMPLE 3. In $\mathbb{Z}[i] = \{a + bi \mid a, b \in \mathbb{Z}, i = \sqrt{-1}\}$ (the ring of Gaussian Integers) the units are $1, -1, i, -i$. So the associates of $2 + 3i$ are $2 + 3i, -2 - 3i, 2i - 3$ and $3 - 2i$.

Definition 10.5. An element p of a commutative ring R with unity is said to be a *prime* if

(i) $p \neq 0$ and p is not a unit.

(ii) For every a, b , in R , $p \mid ab \Rightarrow p \mid a$ or $p \mid b$.

Definition 10.6. An element p in a commutative ring with unity is called *irreducible* if

(i) $p \neq 0$ and p is not a unit.

(ii) Whenever $p = ab$ then one of a and b must be unit; in other words p has no proper factors.

EXAMPLE 4. In $\mathbb{Z}$ every prime number is a prime element and also every prime number is an irreducible element.

EXAMPLE 5. In $\mathbb{Z}[\sqrt{-5}] = \{a + b\sqrt{-5} \mid a, b \in \mathbb{Z}\}$, $\sqrt{-5}$ is

prime element since if $\sqrt{-5} \mid (a+b\sqrt{-5})(c+d\sqrt{-5})$ for some $a, b, c, d \in \mathbb{Z}$, then $(a+b\sqrt{-5})(c+d\sqrt{-5}) = (\sqrt{-5})(x+y\sqrt{-5})$ for some $x, y \in \mathbb{Z}$.

Thus $(a-b\sqrt{-5})(c-d\sqrt{-5}) = (-\sqrt{-5})(x-y\sqrt{-5})$.

So $(a^2+5b^2)(c^2+5d^2) = 5(x^2+5y^2)$

$$\Rightarrow 5 \mid (a^2+5b^2)(c^2+5d^2).$$

But $5 \mid 5b^2c^2+5a^2d^2+25b^2d^2$. Consequently $5 \mid a^2c^2$.

Now 5 is a prime number so either $5 \mid a^2$ or $5 \mid c^2$.

In case $5 \mid a^2$, $5 \mid a$; in second case $5 \mid c$.

Let $5 \mid a$; then $\sqrt{-5} \mid a$ as $5 = (\sqrt{-5})(-\sqrt{-5})$.

Therefore $\sqrt{-5} \mid (a+\sqrt{-5}b)$. Similarly in case $5 \mid c$ we get $\sqrt{-5} \mid (c+\sqrt{-5}d)$. Hence $\sqrt{-5}$ is a prime element of $\mathbb{Z}[\sqrt{-5}]$.

EXAMPLE 6. Again coming back to $\mathbb{Z}[\sqrt{-5}]$ it can be checked that only units in $\mathbb{Z}[\sqrt{-5}]$ are 1 and -1. We claim that 3 is irreducible in $\mathbb{Z}[\sqrt{-5}]$. If it were not, it would be possible to express $3 = (a+b\sqrt{-5})(c+d\sqrt{-5})$ for some $a, b, c, d \in \mathbb{Z}$ where both $a+b\sqrt{-5}$ and $c+d\sqrt{-5}$ are non-units. Then $3 = (a-b\sqrt{-5})(c-d\sqrt{-5}) \Rightarrow 9 = (a^2+5b^2)(c^2+5d^2)$. Thus either $a^2+5b^2=1$ and $c^2+5d^2=9$, or $a^2+5b^2=3$ and $c^2+5d^2=3$ or $a^2+5b^2=9$ and $c^2+5d^2=1$. Since $5 > 1$ the only integral solutions of $a^2+5b^2=1$ are $a = \pm 1, b = 0$. Also $a^2+5b^2=3$ has no integral solution. In case $a^2+5b^2=9, c^2+5d^2=1$, we get $c = \pm 1, d = 0$. Thus $3 = (a+b\sqrt{-5})(c+d\sqrt{-5})$ will either imply $a+b\sqrt{-5} = \pm 1$ or $c+d\sqrt{-5} = \pm 1$. But 1 and -1 are units in $\mathbb{Z}[\sqrt{-5}]$! So 3 is an irreducible element of $\mathbb{Z}[\sqrt{-5}]$. Note that 3 is not prime since $3 \mid 9 = (2+\sqrt{-5})(2-\sqrt{-5})$ but $3 \nmid (2+\sqrt{-5})$ and $3 \nmid (2-\sqrt{-5})$, otherwise $2 \pm \sqrt{-5} = 3(x+y\sqrt{-5})$ for some $x, y \in \mathbb{Z} \Rightarrow 3x = 2$, but $3x = 2$ has no solution in $\mathbb{Z}$.

EXAMPLE 7. In $\mathbb{Z}[i]$, the ring of Gaussian Integers (See Example 3), $1+i$ is an irreducible element. Suppose $1+i = (a+bi)(c+di)$, $a, b, c, d \in \mathbb{Z}$ then $1-i = (a-bi)(c-di) \Rightarrow 2 = (a^2+b^2)(c^2+d^2)$. So either $a^2+b^2=2$ or $a^2+b^2=1$. In case $a^2+b^2=1$ we get $(a+ib)(a-ib) = 1 \Rightarrow a+bi$ is unit. In case $a^2+b^2=2, c^2+d^2=1$ i.e., $(c+di)(c-di) = 1 \Rightarrow c+di$ is unit. Hence $1+i$ is irreducible in $\mathbb{Z}[i]$.

Lemma 10.7. In an integral domain D with unity every prime element is irreducible.

Proof: Let p be a prime element of D , then by definition $p \neq 0$ and p is not a unit. Suppose $p = ab$ for some a and $b \in D$. So $p \mid ab$. But as p is prime either $p \mid a$ or $p \mid b$.

Suppose $p \mid a$ then $a = pk$ for some $k \in D$.

This gives $p = pkb$, so $kb = 1$ and b is a unit. Similarly if $p \mid b$ then a is a unit. Hence p is irreducible. ■

EXAMPLE 8. In $\mathbb{Z}/(6)$, $\bar{2}$ is a prime element but not irreducible

element. Since $2 \mid \bar{a}\bar{b} \Rightarrow 6 \mid (ab-2n) \Rightarrow ab-2n=6k$ for some $k \in \mathbb{Z}$
 $\Rightarrow 2 \mid ab \Rightarrow 2 \mid a$ or $2 \mid b \Rightarrow 2 \mid \bar{a}$ or $2 \mid \bar{b} \Rightarrow \bar{2}$ is a prime element.
 But $\bar{2} = \bar{2} \cdot \bar{4}$ where neither $\bar{2}$ nor $\bar{4}$ is a unit, this gives $\bar{2}$ is not an
 irreducible element. Note that $\mathbb{Z}/(6)$ is not an integral domain.

REMARK. In general, converse of Lemma 10.7 is not true. See Example 6.

Lemma 10.8. Let R be a commutative ring and $a \in R$ then the principal ideal $\langle a \rangle$ is equal to the set $\{ar+na \mid r \in R, n \in \mathbb{Z}\}$.

Proof: Let $X = \{ar+na \mid r \in R, n \in \mathbb{Z}\}$. We shall show that X is the ideal generated by a . Clearly $a = a0 + 1a \in X$ so that X is non-empty. Let $ar+na, as+ma \in X$ where $r, s \in R, n, m \in \mathbb{Z}$. Then $(ar+na) - (as+ma) = a(r-s) + (n-m)a \in X$. Further if $s \in R$ then $(ar+na)s = a(rs) + a(ns) = a(rs+ns) + 0a \in X$. This shows that X is a right ideal of R . As R is commutative, X is also a left ideal of R . Hence X is an ideal of R such that $\{a\} \subseteq X$. Let Y be another ideal of R containing $\{a\}$. Since $a \in Y$ and Y is an ideal of R , $ar \in Y, na \in Y \forall r \in R$ and $\forall n \in \mathbb{Z}$. This implies $ar+na \in Y \forall r \in R$ and $\forall n \in \mathbb{Z}$. Hence $X \subseteq Y$. Consequently $X = \langle a \rangle$. ■

Corollary 10.9. In a commutative ring R with unity $\langle a \rangle = Ra = aR$.

Proof: Let 1 be the unity of R , then we can write $ar+na = a(r+n1) = as$ for some $s \in R$ as $1 \in R \Rightarrow n1 \in R$. Consequently $\langle a \rangle = \{as \mid s \in R\}$ i.e. $\langle a \rangle = aR = Ra$. ■

Definition 10.10. A commutative ring R with unity is called a Principal Ideal Ring (PIR) if each of its ideal is principal. Further if R is a domain then R is called a Principal Ideal Domain (PID).

EXAMPLE 9. Every field is a PID. Now the only ideals of a field F are (0) and F itself. Since for each $x \in F, x = x1$ yields $x \in (1)$, so $F = (1)$. Hence every ideal of F is principal, in other words F is a PID.

However the following theorem provides us examples of PID's which are not fields.

Theorem 10.11. (a) $\mathbb{Z}$ is a PID.

(b) For any field F , the ring of polynomials $F[x]$ is a PID.

Proof: Note that in each part it is sufficient to prove that every non-zero ideal is principal.

(a) Let A be a non-zero ideal of $\mathbb{Z}$, let $a (\neq 0) \in A$. Without loss of generality we may assume that a is positive (since otherwise $-a \in A$ and $-a$ is a positive integer). Let n be the least positive integer in A . We shall show that $A = \langle n \rangle$. Clearly $\langle n \rangle \subseteq A$. Let q be any integer in A . By the division algorithm in $\mathbb{Z}$, there exist integers h and r such that $q = hn + r$ where either $r = 0$ or $0 < r < n$. If $0 < r < n$ then $r = q - hn \in A$ as $q, n \in A$, which contradicts the fact that n is the least

positive integer belonging to A . Thus $r=0$ i.e. $q=hn \Rightarrow q \in \langle n \rangle$. This gives $A \subseteq \langle n \rangle$. Hence $A = \langle n \rangle$ and first assertion of the theorem is established.

(b) Let A be a non-zero ideal of $F[x]$ and let f be a non-zero polynomial in A . Let g be a non-zero polynomial of least degree, which belongs to A , then $\langle g \rangle \subseteq A$. Now if $t \in A$ then by Corollary 9.11, there exist polynomials h and r such that $t = hg + r$, where either $r=0$ or $\deg r < \deg g$. But $r = t - hg \in A$ as $t, g \in A$ and A is an ideal of $F[x]$. So if $r \neq 0$, then $\deg r$ can't be less than $\deg g$ because of our choice of g . In other words $r=0$. This implies $t = hg$ i.e. $t \in \langle g \rangle$. Hence $A = \langle g \rangle$. This completes the theorem. ■

We now show that the converse of Lemma 10.7 holds for PID's.

Lemma 10.12. Every irreducible element of a PID is prime.

Proof: Let p be an irreducible element of a PID R . By definition $p \neq 0$ and p is not a unit. Let $p \mid ab$ and $p \nmid a$; we shall prove that $p \mid b$. Now $\langle p \rangle + \langle b \rangle = \langle d \rangle$ for some $d \in R$ as R is a PID. This gives $\langle b \rangle \subseteq \langle d \rangle \Rightarrow p = dx$ for some $x \in R$. But p is irreducible element so either d is a unit or x is a unit. In case d is a unit, there exist r and $s \in R$ such that $1 = pr + bs \Rightarrow a = apr + abs \Rightarrow p \mid a$ as $p \mid ab$ and $p \mid apr$, which is against our assumption. This enforces x to be a unit. So $d = px^{-1}$. In other words $\langle d \rangle = \langle p \rangle$. Then $\langle p \rangle + \langle b \rangle = \langle p \rangle$ gives $\langle b \rangle \subseteq \langle p \rangle$ i.e. $p \mid b$. Hence the lemma follows. ■

REMARK. The above lemma shows at once that $\mathbb{Z}[\sqrt{-5}]$ is not a PID as a $3 \in \mathbb{Z}[\sqrt{-5}]$ is irreducible but not prime.

Theorem 10.13. Let R be a PID which is not a field. Then any proper ideal A of R is a maximal ideal if and only if it is generated by an irreducible element of R .

Proof: Since R is not a field there exists an element $a (\neq 0) \in R$ such that a is not a unit, then $(0) \subsetneq \langle a \rangle \subsetneq R$. So $\langle 0 \rangle$ is not a maximal ideal. Now let A be a maximal ideal of R ; then as seen above $A \neq (0)$. Now $A = \langle p \rangle$ for some $p (\neq 0) \in R$ as R is a PID. Further, $A \neq R \Rightarrow p$ is not a unit. We show that p is an irreducible element. Let $p = ab$ for some $a, b \in R, p \in \langle a \rangle \Rightarrow A \subseteq \langle a \rangle$. As A is a maximal ideal, either $A = \langle a \rangle$ or $\langle a \rangle = R$. In case $\langle a \rangle = R, 1 = ac$ for some $c \in R \Rightarrow a$ is unit. In case $A = \langle a \rangle, a \in \langle p \rangle \Rightarrow a = pq$ for some $q \in R$, we get $p = pqb$ which in turn gives $qb = 1$ i.e. b is a unit, hence p is irreducible.

Conversely let $A = \langle p \rangle$ where p is an irreducible element. Since p is a non-unit, $A \neq R$. Let $J \supsetneq A$ be an ideal of R . As R is a PID, $J = \langle x \rangle$ for some $x \in R$ and $x \notin A$ (otherwise $J = A$). Now $\langle p \rangle \subsetneq \langle x \rangle \Rightarrow x \mid p$ i.e. $p = xy$ for some $y \in R$. As p is irreducible element either x is a unit or y is a unit. If y is a unit then

$x=py^{-1} \in A$, a contradiction. This forces x to be a unit. In other words $J=\langle x \rangle=R$. Hence A is a maximal ideal of R . ■

EXAMPLE 10. In $\mathbf{Z}$, an ideal $\langle p \rangle$ is maximal if and only if p is a prime number. This follows from the fact that in $\mathbf{Z}$ each irreducible element is a prime number and vice versa.

EXAMPLE 11. In $\mathbf{Q}[x]$, the ideal $\langle x+2 \rangle$ is maximal. Note that $\mathbf{Q}[x]$ is a PID (Theorem 10.11) and $x+2$ is an irreducible element in $\mathbf{Q}[x]$, because if $x+2=fg$, then $\deg f + \deg g = 1 \Rightarrow$ either $\deg f=0$ or $\deg g=0$, let $\deg f=0$ then $\deg g=1$; let $f=a_0$ where $a_0 \neq 0 \in \mathbf{Q}$ and let $g=b_0+b_1x$ where $b_0, b_1 \in \mathbf{Q}$ and $b_1 \neq 0$. Then $x+2 = a_0(b_0+b_1x) \Rightarrow a_0b_0=2$ and $a_0b_1=1$. So a_0 is a unit i.e. f is a unit. Similarly $\deg g=0 \Rightarrow g$ is a unit, hence $x+2$ is an irreducible element. Then by the above theorem $\langle x+2 \rangle$ is a maximal ideal.

Definition 10.14. (HCF). Let R be a commutative ring. Given two non-zero elements a and b of R , a non-zero element c of R is said to be a *Highest Common Factor (HCF)* of a and b in R if

- (i) $c \mid a$ and $c \mid b$ in R
- and (ii) for any $d \neq 0 \in R$ if $d \mid a$ and $d \mid b$ in R then $d \mid c$ in R .

Definition 10.15. (LCM). A non-zero element d of a commutative ring R is said to be a *Least Common Multiple (LCM)* of two non zero elements a and b of R , in R , if

- (i) $a \mid d$ and $b \mid d$
- and (ii) for any $x (\neq 0)$ in R if $a \mid x$ and $b \mid x$ in R , then $d \mid x$ in R .

Both definitions can be extended in a natural way to any finite number of non-zero elements of R .

EXAMPLE 12. In $\mathbf{Z}$, 3 is an *HCF* of 6 and 9. Note that $6=3 \cdot 2$ and $9=3 \cdot 3$ so $3 \mid 6$ and $3 \mid 9$. Further let $a (\neq 0)$ be an integer such that $a \mid 6$ and $a \mid 9$ then $a \mid (9-6)$ i.e. $a \mid 3$. Hence 3 satisfies conditions (i) and (ii) of the definition of *HCF*. Similarly it can be shown that -3 is also an *HCF* of 6 and 9. Again 18 is an *LCM* of 6 and 9. Now $18=6 \cdot 3$ and $18=9 \cdot 2$ imply $6 \mid 18$ and $9 \mid 18$. Further let $x (\neq 0)$ an integer such that $6 \mid x$ and $9 \mid x$. Now $2 \mid 6 \Rightarrow 2 \mid x$ and $3 \mid 9 \Rightarrow 3 \mid x \Rightarrow x=6l$ for some integer l . This gives $18k-4.18l=x \Rightarrow 18 \mid x$. Hence 18 is *LCM* of 6 and 9. In the same way it can be proved that -18 is also an *LCM* of 6 and 9.

EXAMPLE 13. In the ring $\mathbf{Z}/(12)$, the elements are $\bar{0}, \bar{1}, \bar{2}, \bar{3}, \bar{4}, \bar{5}, \bar{6}, \bar{7}, \bar{8}, \bar{9}, \bar{10}$ and $\bar{11}$. *HCF* of $\bar{6}$ and $\bar{8}$ is $\bar{2}$. Notice that $\bar{6}=\bar{3} \cdot \bar{2}$ and $\bar{8}=\bar{4} \cdot \bar{2}$. Further if $\bar{x} \neq \bar{0}$ in $\mathbf{Z}/(12)$ be such that $\bar{x} \mid \bar{6}$ and $\bar{x} \mid \bar{8}$ then $\bar{x} \mid (\bar{8}-\bar{6})$ i.e. $\bar{x} \mid \bar{2}$. Hence $\bar{2}$ is an *HCF* of $\bar{6}$ and $\bar{8}$. Note that $\bar{6}=\bar{10} \cdot \bar{3}$ and $\bar{8}=\bar{10} \cdot \bar{2}$. So $\bar{10} \mid \bar{6}$ and $\bar{10} \mid \bar{8}$. Further if $\bar{y} (\neq \bar{0}) \in \mathbf{Z}/(12)$ be such that $\bar{y} \mid \bar{6}$ and $\bar{y} \mid \bar{8}$ then $\bar{y} \mid (\bar{2} \cdot \bar{8} - \bar{6})$ i.e. $\bar{y} \mid \bar{10}$. Hence $\bar{10}$ is

also an *HCF* of $\bar{6}$ and $\bar{8}$. We claim that $\bar{6}$ and $\bar{8}$ have no *LCM* in $\mathbf{Z}/(12)$. Suppose that $\bar{x}$ is an *LCM* of $\bar{6}$ and $\bar{8}$ then $\bar{6} \mid \bar{x}$ and $\bar{8} \mid \bar{x}$ so $\bar{x} = \bar{6}\bar{n}$ for some $\bar{n} \in \mathbf{Z}/(12)$. Then $\bar{x} = \bar{0}$ or $\bar{6}$, this implies $\bar{6} = \bar{8}\bar{q}$ for some $\bar{q} \in \mathbf{Z}/(12)$ as *LCM* is never zero. But $\bar{8}\bar{m} = \bar{0}$, $\bar{4}$ or $\bar{8} \nmid \bar{m} \in \mathbf{Z}/(12)$, hence a contradiction. Consequently $\bar{6}$ and $\bar{8}$ have no *LCM* in $\mathbf{Z}/(12)$. This shows that elements may possess an *HCF* but not an *LCM*.

EXAMPLE 14. In $\mathbf{Z}[x]$, the ring of polynomials over $\mathbf{Z}$, $x+2$ is an *HCF* of x^2+2x and $2x+4$ while $2x^2+4x$ is their *LCM*. Clearly $(x+2) \mid (x^2+2x)$ and $(x+2) \mid (2x+4)$. Let $f(\neq 0) \in \mathbf{Z}[x]$ be such that $f \mid (x^2+2x)$ and $f \mid (2x+4)$. Then $2x+4 = fg$ for some $g \in \mathbf{Z}[x]$. Since $\mathbf{Z}$ is an integral domain, $\deg f + \deg g = \deg(2x+4) = 1$. So either $\deg f = 0$ or $\deg g = 0$. In case $\deg f = 0$, $f = a_0$, $a_0 \neq 0$, then $f \mid (x^2+2x) \Rightarrow x^2+2x = a_0(b_0+b_1x+b_2x^2)$, $b_2 \neq 0$ as $\deg(x^2+2x) = 2$; this gives that $a_0b_2 = 1 \Rightarrow a_0$ is unit of $\mathbf{Z} \Rightarrow a_0 = 1$ or -1 . Then $f = 1$ or -1 and so $f \mid (x+2)$.

In case $\deg f = 1$, let $f = a_0 + a_1x$, $a_1 \neq 0$.

Now $f \mid (2x+4) \Rightarrow 2x+4 = (a_0+a_1x)c_0$, $c_0 \neq 0 \Rightarrow a_0c_0 = 4 \Rightarrow a_0 \neq 0$. Again $f \mid (x^2+2x) \Rightarrow x^2+2x = (a_0+a_1x)(d_0+d_1x)$, $d_1 \neq 0$ as $\deg(x^2+2x) = 2$; $\Rightarrow a_1d_0 = 0$ and $a_0 \neq 0 \Rightarrow d_0 = 0$. Further $a_1d_1 = 1 \Rightarrow a_1$ is a unit so $a_1 = \pm 1$; This gives $d_1 = \pm 1$. Finally $a_0d_1 = 2 \Rightarrow a_0 = 2$ in case $d_1 = 1$ and $a_0 = -2$ in case $d_1 = -1$, hence $f = x+2$ or $-x-2$. In both the cases $f \mid (x+2)$.

Thus $x+2$ is an *HCF* of x^2+2x and $2x+4$. We leave the verification of the proof of the statement that $2x^2+4x$ is an *LCM* of x^2+2x and $2x+4$ to the readers.

One notices in the above examples that a pair of non-zero elements a and b of a commutative ring may have more than one *HCF* and the existence of *HCF* does not guarantee the existence of *LCM*. Similar is the remark about *LCM* and its existence *vis-a-vis* *HCF*. Also one can't fail to notice that if there are more than one *HCF* or *LCM* in an integral domain with unity then all these are associates. We establish this in the following.

Theorem 10.16. Let D be an integral domain with unity. If a and b are two non-zero elements of D such that their *HCF* exist then any two *HCF*'s of a and b are associates. Similarly if a and b have *LCM* then any two *LCM*'s of a and b are associates.

Proof: We shall give the proof for the case of *HCF* and leave the second case for the readers to prove.

Let c and d be two *HCF*'s of a and b then since $c \mid a$ and $c \mid b$ and d is *HCF* of a and b , $c \mid d$. Similarly $d \mid c$. Hence c and d are associates. [Theorem 10.4(ii)]. ■

NOTATION. *HCF* of a and b will be denoted by (a, b) and *LCM* of a and b will be denoted by $[a, b]$.

Lastly we give an example to show that a pair of non-zero elements in an integral domain with unity may not possess an *HCF*.

EXAMPLE 15. Consider the ring $R = \mathbb{Z}[\sqrt{-5}] = \{a + b\sqrt{-5} \mid a, b \in \mathbb{Z}\}$.

Then $9, -3 + 3\sqrt{-5} \in \mathbb{Z}[\sqrt{-5}]$, $3 \mid 9$ and $3 \mid -3 + 3\sqrt{-5}$.

Firstly we show the only proper factors of 9 are $3, 3 : -3, -3; 2 + \sqrt{-5}, 2 - \sqrt{-5}; -2 - \sqrt{-5}, -2 + \sqrt{-5}$.

Suppose $9 = (a + b\sqrt{-5})(c + d\sqrt{-5})$ where $a + b\sqrt{-5}$ and $c + d\sqrt{-5}$ are non-units. This gives $81 = N[(a + b\sqrt{-5})(c + d\sqrt{-5})]$ where $N(x + y\sqrt{-5}) = x^2 + 5y^2$; i.e. $81 = (a^2 + 5b^2)(c^2 + 5d^2)$.

So $a^2 + 5b^2 = 1, 3, 9, 27$ or 81 .

In case $a^2 + 5b^2 = 1$, $a = \pm 1, b = 0$ then $a + b\sqrt{-5}$ is a unit which is a contradiction.

$a^2 + 5b^2 = 3$ has no integral solution. (Prove !)

$a^2 + 5b^2 = 9$ has $a = \pm 2, b = \pm 1$ or $a = \pm 3, b = 0$ as the integral solutions. Again $c^2 + 5d^2 = 9$ gives $c = \pm 2, d = \pm 1$ or $c = \pm 3, d = 0$. Thus we get $\pm 3, \pm 2 \pm \sqrt{-5}$ as six proper factors of 9.

Again $a^2 + 5b^2 = 27$ has no integral solution.

Finally $a^2 + 5b^2 = 81 \Rightarrow c^2 + 5d^2 = 1$ which in turn will imply $c + d\sqrt{-5}$ is a unit which is absurd. Hence our assertion follows.

Now none of the factors $2 + \sqrt{-5}, 2 - \sqrt{-5}, -2 - \sqrt{-5}$ or $-2 + \sqrt{-5}$ can be *HCF* of 9 and $-3 + 3\sqrt{-5}$ otherwise $3 \mid (2 + \sqrt{-5})$ or $3 \mid (2 - \sqrt{-5})$ which is impossible as shown in Example 6.

Further 9 itself can't be an *HCF* of 9 and $-3 + 3\sqrt{-5}$ as $9 \mid (-3 + 3\sqrt{-5}) \Rightarrow -3 + 3\sqrt{-5} = 9(u + v\sqrt{-5})$ for some $u, v \in \mathbb{Z} \Rightarrow 9u = -3$ has a solution in integers which is absurd.

Hence if there is at all an *HCF* of 9 and $-3 + 3\sqrt{-5}$ it must be 3, (Note that an associate of *HCF* is again an *HCF*) Let it be so.

Now $2 + \sqrt{-5}$ is a factor of 9 and also.

$(-3 + 3\sqrt{-5}) = (2 + \sqrt{-5})(1 + \sqrt{-5}) \Rightarrow (2 + \sqrt{-5}) \mid (-3 + 3\sqrt{-5})$.

So $(2 + \sqrt{-5}) \mid 3$ as 3 is *HCF* of 9 and $-3 + 3\sqrt{-5}$ by our assumption. But in that case $3 = (2 + \sqrt{-5})(p + q\sqrt{-5})$ and $3 = (2 - \sqrt{-5})(p - q\sqrt{-5}) \Rightarrow 9 = 9(p^2 + 5q^2) \Rightarrow p^2 + 5q^2 = 1 \Rightarrow p = \pm 1, q = 0 \Rightarrow 3 = \pm(2 + \sqrt{-5})$ which is absurd. Hence 9 and $-3 + 3\sqrt{-5}$ have no *HCF*.

Theorem 10.17. If R is a PID then every pair of non-zero elements a and b of R has an *HCF* and an *LCM*. Further if $d = (a, b)$ then $d = ax + by$ for some $x, y \in R$.

Proof: Consider $\langle a \rangle + \langle b \rangle$. Since R is a PID, $\langle a \rangle + \langle b \rangle = \langle c \rangle$ for some $c \in R$. Thus $\langle a \rangle \subseteq \langle c \rangle$ and $\langle b \rangle \subseteq \langle c \rangle \Rightarrow$

$c \mid a$ and $c \mid b$. Let $d(\neq 0) \in R$ be such that $d \mid a$ and $d \mid b$ then $\langle a \rangle \subseteq \langle d \rangle$ and $\langle b \rangle \subseteq \langle d \rangle \Rightarrow \langle a \rangle + \langle b \rangle \subseteq \langle d \rangle \Rightarrow \langle c \rangle \subseteq \langle d \rangle \Rightarrow d \mid c$.

Hence c is an HCF of a and b .

Now $c \in \langle a \rangle + \langle b \rangle \Rightarrow c = ax + by$ for some $x, y \in R$.

Again consider $\langle a \rangle \cap \langle b \rangle$, since R is a PID, $\langle a \rangle \cap \langle b \rangle = \langle l \rangle$ for some $l \in R$. Now $\langle l \rangle \subseteq \langle a \rangle$ and $\langle l \rangle \subseteq \langle b \rangle \Rightarrow a \mid l$ and $b \mid l$. Let $m(\neq 0) \in R$ be such that $a \mid m$ and $b \mid m \Rightarrow \langle m \rangle \subseteq \langle a \rangle$ and $\langle m \rangle \subseteq \langle b \rangle \Rightarrow \langle m \rangle \subseteq \langle a \rangle \cap \langle b \rangle = \langle l \rangle \Rightarrow l \mid m$.

Hence l is an LCM of a and b . ■

Corollary 10.18. Let R be a PID then any n non-zero elements $a_1, a_2, \dots, a_n \in R$ have an HCF and an LCM.

Proof: Apply induction on n . ■

Definition 10.19. Two non-zero elements a and b of a commutative ring with unity are said to be co-prime or relatively prime to each other, in case their HCF is a unit.

Corollary 10.20. (Bezout's Identity). Let R be a PID. Two non-zero elements a and b of R are co-prime if and only if there exist $x, y \in R$ such that $ax + by = 1$.

Proof: If a and b are co-prime then 1 being associate of every unit in R , is HCF of a and b , so there exist $x, y \in R$ such that $ax + by = 1$ (Theorem 10.17). Conversely let there exist $x, y \in R$ such that $ax + by = 1$. If d is an HCF of a and b then $d \mid a$ and $d \mid b \Rightarrow d \mid (ax + by) \Rightarrow d \mid 1$. Hence $1 = dc$ for some $c \in R$. In other words d is a unit. Consequently a and b are co-prime. ■

Corollary 10.21. (Generalized Bezout's Identity). If R is a PID, then n non-zero elements $a_1, a_2, \dots, a_n$ of R are relatively prime (i.e. their HCF is a unit) if and only if there exist $x_1, x_2, \dots, x_n \in R$ such that $a_1x_1 + a_2x_2 + \dots + a_nx_n = 1$.

Proof is similar to that of Corollary 10.20. ■

PROBLEMS

1. Show that an associate of an HCF of two non-zero elements a and b of a commutative ring R with unity is again an HCF of a and b . Prove the same result for LCM.

2. If $a_1, a_2, \dots, a_n$ and r are non-zero elements of an integral domain R show that if HCF of $ra_1, ra_2, \dots, ra_n$ exists then HCF of $a_1, a_2, \dots, a_n$ exists and $(ra_1, ra_2, \dots, ra_n) = r(a_1, a_2, \dots, a_n)$.

Also prove that if LCM of $a_1, a_2, \dots, a_n$ exists then LCM of $ra_1, ra_2, \dots, ra_n$ exists and $[ra_1, ra_2, \dots, ra_n] = r[a_1, a_2, \dots, a_n]$.

[Hint. Let $c = (ra_1, ra_2)$. Now $r \mid ra_1$ and $r \mid ra_2 \Rightarrow r \mid c \Rightarrow c = rp$

for some $d \in R$. Now clearly $d \mid a_1$ and $d \mid a_2$ as $c \mid ra_1$ and $c \mid ra_2$. Further if $x \mid a_1$ and $x \mid a_2$ then $rx \mid ra_1$ and $rx \mid ra_2 \Rightarrow rx \mid c \Rightarrow x \mid d$. Again if $l = [a_1, a_2]$ then clearly $ra_1 \mid rl$ and $ra_2 \mid rl$. Let $ra_1 \mid x$ and $ra_2 \mid x$ then $r \mid x \Rightarrow x = rs$ for some $s \in R$. This gives $a_1 \mid s$ and $a_2 \mid s \Rightarrow l \mid s \Rightarrow rl \mid x$

3. Prove that in an integral domain with unity, every pair of non-zero elements have HCF if and only if every pair of non-zero elements have LCM.

Prove that if R is a PID and a, b are two non-zero elements of R then $[a, b](a, b) = abu$ for some unit $u \in R$.

5. Prove that an integral domain R with unity is a field if and only if $R[x]$ is a PID. Deduce that $\mathbf{Z}[x]$ is not a PID.

6. Let R be a commutative ring with unity. Let $a, e \in R$ and let $e^2 = e$. Prove:

(i) If $\langle a \rangle = \langle e \rangle$ then a and e are associates.

(ii) If for some positive integer m , a^m and e are associates then for all integers $n \geq m$, a^n and e are also associates.

7. Find HCF and LCM of the following if they exist.

(i) $\bar{9}$ and $\bar{18}$ in $\mathbf{Z}/(20)$.

(ii) $x^2 + 3x$ and $3x + 9$ in $\mathbf{Z}[x]$.

(iii) $x^2 + 5x + 6$ and $x^2 + 7x + 12$ in $\mathbf{Z}[x]$.

(iv) $\bar{7}$ and $\bar{6}$ in $\mathbf{Z}/(14)$.

(v) $3x^5 - 8x^4 - x^2 + 6x - 2$ and $x^9 - 3x^8 + x^7 - 2x^2 + 6x - 2$ in $\mathbf{Z}[x]$.

Ans. (i) HCF $\bar{9}$; LCM $\bar{18}$.

(ii) HCF $x + 3$; LCM $3x^2 + 9x$.

(iii) HCF $x + 3$; LCM $x^3 + 9x^2 + 26x + 24$.

(iv) HCF $\bar{1}$; LCM does not exist.

(v) HCF $x^2 - 3x + 1$; LCM $(x^2 - 3x + 1)(3x^3 + x^2 - 2)(x^7 - 2)$.

[Recall that every associate of HCF is an HCF and every associate of LCM is an LCM].

8. If A, B, C are non-zero ideals of a ring R , show that if R is a PID then

(a) $A(B \cap C) = AB \cap AC$.

(b) $A = \langle a \rangle$ and $B = \langle b \rangle \Rightarrow AB = \langle ab \rangle$.

9. Show that if R is a PID, $AB = A \cap B$ if and only if A and B are co-maximal; provided A and B are non-zero ideals of R .

10. Let p be a prime element of a commutative ring R with unity. Show that if $p \mid a_1 a_2 \dots a_n$ then there must exist at least one i , $1 \leq i \leq n$ such that $p \mid a_i$.

11. Let a be a non-zero element of a PID and let $a \mid bc$ be such that $(a, b) = 1$. Show that $a \mid c$.

12. If in a PID, three non-zero elements a, b and c are such that

$a \mid c$ and $b \mid c$ and further $(a, b)=1$, show that $ab \mid c$.

13. Show that in $\mathbb{Z}[x]$, $\langle x \rangle + \langle 2 \rangle$ is not a principal ideal.

14. Show that in E , the ring of even integers, every ideal is principal.

15. Let R be a PID and let a, b, c be three non-zero elements of R , prove that

(i) $(a, b)=1$ and $(a, c)=1 \Rightarrow (a, bc)=1$

(ii) $(a, (b, c))=((a, b), c)$.

(iii) $(a, 1)=1$.

16. Show that in any commutative ring with unity, an associate of prime (irreducible) element is prime (irreducible).

17. If a prime element p of an integral domain with unity, divides another prime element q , prove that p and q are associates.

2. Euclidean Domains

We know that if a and b are two integers ($b \neq 0$), then there exist integers c and d such that $a=bc+d$ where $d=0$ or $0 < d < |b|$. In this section we intend to generalize this concept.

Definition 10.22. A non zero integral domain R is called a *Euclidean domain*, if there exists a function δ from $R - \{0\}$ to $\mathbb{Z}$ such that

(1) $\delta(a) \geq 0 \forall a \in R - \{0\}$.

(2) $\forall a, b \in R - \{0\}, \delta(ab) \geq \delta(a)$.

(3) $\forall a \in R$ and $b \in R - \{0\}$ there exist $q, r \in R$ such that $a=qb+r$ with either $r=0$ or $\delta(r) < \delta(b)$.

REMARK. Condition (3) is called a *Euclidean Algorithm* and the function δ is called a *Euclidean Valuation* on R . Note also that in condition (2) we can also write $\delta(ab) \geq \delta(b)$.

EXAMPLE 16. The ring $\mathbb{Z}$ of integers is a Euclidean Domain.

Here we take $\delta(a) = |a| \forall a \in \mathbb{Z} - \{0\}$. Clearly δ satisfies all the conditions (1), (2) and (3).

EXAMPLE 17. Every field F is a Euclidean Domain. We take $\delta(a) = 1 \forall a \in F - \{0\}$. For all non-zero $a, b \in F, ab \neq 0$ so $\delta(ab) = 1 = \delta(a)$. Further if $a \in F$ and b is any non zero element of F , we can write $a = (ab^{-1})b + 0$. Hence (3) is also true.

EXAMPLE 18. The ring $\mathbb{Z}[i] = \{a+bi \mid a, b \in \mathbb{Z}, i = \sqrt{-1}\}$ of Gaussian Integers is a Euclidean Domain. We define $\delta : \mathbb{Z}[i] \rightarrow \mathbb{Z}$ by $\delta(a+bi) = a^2 + b^2$ for all $a+bi$. Clearly (1) is satisfied and for non-zeros $a+bi, c+di \in \mathbb{Z}[i]$, $\delta[(a+bi)(c+di)] = \delta[(ac-bd) + (ad+bc)i] = (ac-bd)^2 + (ad+bc)^2 = (a^2+b^2)(c^2+d^2) \geq a^2+b^2$ as $c^2+d^2 \geq 1$, so $\delta[(a+bi)(c+di)] \geq \delta(a+bi)$. Thus, condition (2) holds.

Now to prove (3) we proceed as under:

Let $a+bi \in \mathbb{Z}[i]$ and $c+di (\neq 0)$ in $\mathbb{Z}[i]$. (In other words c and d

are not simultaneously zero). Consider $\frac{a+bi}{c+di}$, this is an element of complex number field. Now $\frac{a+bi}{c+di} = p+qi$ where $p = \frac{ac+bd}{c^2+d^2}$ and $q = \frac{bc-ad}{c^2+d^2}$ are rational numbers.

There exist integers m and n such that $|p-m| \leq \frac{1}{2}$ and $|q-n| \leq \frac{1}{2}$. {Note that $p=[p]+\theta$ where $[p]$ is integral part of p and θ is a fractional part. If $0 \leq \theta \leq \frac{1}{2}$ we take $m=[p]$ and if $\frac{1}{2} < \theta < 1$ we take $m=[p]+1$.}

Let $p-m=\alpha$ and $q-n=\beta$ then $|\alpha| \leq \frac{1}{2}$ and $|\beta| \leq \frac{1}{2}$. Now $a+bi=(p+qi)(c+di)$

$$\begin{aligned} &= (c+di)[m+\alpha+(n+\beta)i] = (c+di)[(m+ni)+(\alpha+\beta i)] \\ &= (m+ni)(c+di) + (c+di)(\alpha+\beta i) \\ &= (m+ni)(c+di) + r = q'(c+di) + r \end{aligned}$$

where $q' = m+ni \in \mathbf{Z}[i]$ and $r = (c+di)(\alpha+\beta i)$.

Now $a+bi, c+di, m+ni \in \mathbf{Z}[i]$ as $a, b, c, d, m, n \in \mathbf{Z}$, so $r = (a+bi) - (m+ni)(c+di) \in \mathbf{Z}[i]$.

Also if $r \neq 0$ then $\delta(r) = (c^2+d^2)(\alpha^2+\beta^2) \leq (c^2+d^2)(\frac{1}{4}+\frac{1}{4})$
 $\Rightarrow \delta(r) \leq \frac{1}{2}(c^2+d^2) < (c^2+d^2) \Rightarrow \delta(r) < \delta(c+di)$.

This gives $(a+bi) = q'(c+di) + r$ where $r=0$ or $\delta(r) < \delta(c+di)$. Hence $\mathbf{Z}[i]$ is a Euclidean Domain.

EXAMPLE 19. If F is a field then the ring of polynomials $F[x]$ over F is a Euclidean Domain. $F[x]$ is an integral domain as F is an integral domain. We know that degree of every non-zero polynomial is a non-negative integer so we define $\delta: F[x] - \{0\} \rightarrow \mathbf{Z}$ by $\delta(f) = \deg f \forall f \in F[x] - \{0\}$. Since F is an integral domain $\forall f, g \in F[x] - \{0\}$, $\delta(fg) = \deg fg = \deg f + \deg g \geq \deg f = \delta(f)$. So condition (2) for a Euclidean Domain holds. Finally by division algorithm in $F[x]$, for $f \in F[x], g(\neq 0) \in F[x]$ there exist polynomials h and $r \in F[x]$ such that $f = hg + r$ where either $r=0$ or $\deg r < \deg g$ i.e. either $r=0$ or $\delta(r) < \delta(g)$. Thus (3) also holds for $F[x]$. Hence $F[x]$ is a Euclidean Domain.

Theorem 10.23. Every Euclidean Domain is a PID.

Proof: Let R be a Euclidean Domain. Let $A \neq (0)$ be an ideal of R , then there exists a non-zero element $a \in A$. Let $M = \{\delta(x) \mid x(\neq 0) \in A\}$. Clearly $\delta(a) \in M$, so M is a non-empty set of non-negative integers, so by well ordering property of non-negative integers M has a least element. Let $b(\neq 0) \in A$ be such that $\delta(b)$ is the least element of M . We claim that $A = \langle b \rangle$. Clearly $\langle b \rangle \subseteq A$. Now if $y \in A$ then as R is a Euclidean Domain there exist $q, r \in R$ such that $y = qb + r$ where either $r=0$ or $\delta(r) < \delta(b)$.

... (1)

Now $r = y - qb \in A$ as $y \in A$, $b \in A$ and A is an ideal of R , so if $r \neq 0$, $\delta(r) \geq \delta(b)$ because of choice of b , which is a contradiction to (1). This forces r to be 0.

Hence $y = qb \in \langle b \rangle$. So that $A \subseteq \langle b \rangle$; consequently $A = \langle b \rangle$. Hence every ideal of R is a principal ideal. Incidentally we have shown that $A = bR$. Finally we show that R has unity.

Since R itself is an ideal of R , there exists $c \in R$ such that $R = cR$. Now since $c \in R$, $c = ce$ for some $e \in R$.

Then for any $x \in R$, $x = cy$ for some $y \in R \Rightarrow xe = cye = cey = cy = x$. Hence e is the unity of R . Therefore R is a PID. ■

REMARK 1. Note that we have shown that every Euclidean Domain possesses unity.

REMARK 2. Converse of Theorem 10.23 is not true. The ring $R = \left\{ a + b \left(\frac{1 + \sqrt{-19}}{2} \right) \mid a, b \in \mathbb{Z} \right\}$ is a PID but not a Euclidean Domain. See Appendix, page 506.

PROBLEMS

1. Show that the following rings are Euclidean Domains.

- (i) $\mathbb{Z}[\sqrt{2}] = \{a + b\sqrt{2} \mid a, b \in \mathbb{Z}\}$
- (ii) $\mathbb{Z}[\sqrt{-2}] = \{a + b\sqrt{-2} \mid a, b \in \mathbb{Z}\}$
- (iii) $\mathbb{Z}[\sqrt{3}] = \{a + b\sqrt{3} \mid a, b \in \mathbb{Z}\}$.

2. Show that $\mathbb{Z}[\sqrt{-5}] = \{a + b\sqrt{-5} \mid a, b \in \mathbb{Z}\}$ is not a Euclidean Domain.

[Hint. Use the fact that a Euclidean Domain is a PID.]

3. Without using Theorem 10.17, prove that every pair of non-zero elements a and b of a Euclidean Domain D have HCF.

[Hint Consider the set $M = \{\lambda a + \mu b \mid \lambda, \mu \in D\}$ and show that it is a principal ideal.]

4. Prove that a non-zero element a of a Euclidean Domain D is a unit if and only if $\delta(a) = \delta(1)$.

5. Let D be a Euclidean Domain, prove that

- (i) $\delta(a) = \delta(-a) \forall a (\neq 0) \in D$.
- (ii) If for a non zero element a of D , $\delta(a) = 0$, then a is a unit.

6. (a) If a, b are non-zero elements of a Euclidean Domain, prove that:—(1) If a and b are associates then $\delta(a) = \delta(b)$.

(2) If $a \mid b$ and $\delta(a) = \delta(b)$ then a and b are associates.

(3) $\delta(ab) > \delta(b)$ if and only if a is not a unit.

(b) Give an example of two elements a, b , such that $\delta(a) = \delta(b)$ but a and b are not associates.

7. Prove the following in a Euclidean Domain D .

(i) If $a(\neq 0) \in D$ is a non-unit then $a = p_1 p_2 \dots p_n$, a product of prime elements.

(ii) If $a = p_1 p_2 \dots p_n = q_1 q_2 \dots q_m$ are two expressions of a as a product of primes, prove that $n = m$ and there exists a permutation σ of $\{1, 2, 3, \dots, n\}$ such that $p_i \sim q_{\sigma(i)}$ for $i = 1, 2, \dots, n$.

[Hint. Use induction on $\delta(a)$].

8. If D is a Euclidean Domain prove that for every integer n such that $\delta(1) + n \geq 0$, the function $f_n : D - \{0\} \rightarrow \mathbf{Z}$ defined by $f_n(a) = \delta(a) + n$ is a Euclidean valuation on D .

9. Show that given a positive integer n , if we define $\delta(a) = |a|^n$ for all $a(\neq 0)$ in $\mathbf{Z}$, then δ is a Euclidean valuation on $\mathbf{Z}$.

10. Show that for every proper ideal A of the ring $\mathbf{Z}[i]$, the quotient ring $\mathbf{Z}[i]/A$ is a finite ring.

11. Let R be a Euclidean Domain and let a, b be two non-zero elements of R . We know that there exist q, r in R such that $a = bq + r$ with either $r = 0$ or $\delta(r) < \delta(b)$. Show that for all pairs of non-zero elements a, b , of R , q and r are unique if and only if for all non-zero $x, y \in R$, $x + y \neq 0$ implies $\delta(x + y) \leq \max[\delta(x), \delta(y)]$.

[Hint. If $\delta(x + y) > \max[\delta(x), \delta(y)]$, we can write $x = 1(x + y) - y$ and $x = 0(x + y) + x$.]

12. Let R be a Euclidean Domain with δ as its Euclidean valuation and let $F = \{a \in R \mid \delta(a) = \delta(1)\} \cup \{0\}$. If the set F is closed under addition, show that F is a subfield of R .

3 Unique Factorization Domains

The Fundamental Theorem of Arithmetic asserts that every integer greater than 1 can be expressed as a product of finite number of primes and this expression is unique except for order. The motivation behind this section is to study those rings which possess property analogous to the Fundamental Theorem of Arithmetic.

Definition 10.24. An integral domain R with unity is called a *Unique Factorization Domain (UFD)* or a *Gaussian domain* if

(1) Each non-zero non-unit a in R is expressible as a product of finite number of irreducible elements of R .

(2) If $a = p_1 p_2 \dots p_r$ and $a = q_1 q_2 \dots q_s$ are two expressions (or decompositions) of a as a product of irreducible elements, then $r = s$ and there exists a permutation σ of $\{1, 2, 3, \dots, r\}$ such that p_i is an associate of $q_{\sigma(i)}$ for every $i = 1, 2, \dots, r$. [i.e. there exists a 1-1 correspondence between p_i 's and q_j 's such that the corresponding elements are associates].

Thus, the conditions (1) and (2) give that in a UFD every non-zero non-unit element is expressible as a product of irreducible elements

and this expression is unique to within associates and orders. We can state briefly that any two expressions of non-zero non-unit as a product of irreducible elements are *equivalent*.

EXAMPLE 20. Ring $\mathbf{Z}$ is a UFD, since in $\mathbf{Z}$ every integer different from -1 , 0 and 1 is expressible as a product of finite number of prime numbers and this expression is unique except for order and sign (Notice that in $\mathbf{Z}$, the prime numbers are nothing but irreducible elements and their negatives are also irreducible elements *i.e.* -3 and 3 both are irreducible elements.)

EXAMPLE 21. Every field F is a UFD. Since if it were not, we would have found a non-zero non-unit in F which is not expressible as a product of irreducible elements. But in a field there is no non-zero non-unit. Hence F is a UFD.

Lemma 10.25. In a UFD every irreducible element is prime.

Proof. Let R be a UFD and p be an irreducible element of R . Let $p \mid ab$ for some a, b in R . Then $ab = cp$ for some $c \in R$. Since $p \neq 0$ and p is not a unit, at least one of a and b is a non-unit. Suppose a is a unit then a^{-1} exists and $b = cpa^{-1}$ gives $p \mid b$. Similarly if b is unit then $p \mid a$. Suppose a and b both are non-units. Then we can write

$$a = p_1 p_2 \dots p_n, b = q_1 q_2 \dots q_m$$

as products of irreducible elements. If c is a unit then ab is an associate of p , so irreducible. This in turn implies that either a is unit or b is unit which is a contradiction. Hence c is a non-unit. So we can write $c = r_1 r_2 \dots r_s$, a product of irreducible elements.

As a consequence we get

$$p_1 p_2 \dots p_n q_1 q_2 \dots q_m = r_1 r_2 \dots r_s p$$

This equation gives that p is either associate of a p_i or an associate of q_j . So either $p \mid a$ or $p \mid b$. Hence in any case $p \mid a$ or $p \mid b$. Thus p is a prime element. ■

REMARK. We have seen in Example 6, that 3 is an irreducible element of $\mathbf{Z}[\sqrt{-5}]$ but not a prime element. In view of the result proved above we get that $\mathbf{Z}[\sqrt{-5}]$ is not a UFD.

Lemma 10.26. In any ring R , the union of any ascending chain of ideals

$$A_1 \subseteq A_2 \subseteq A_3 \subseteq \dots \subseteq A_n \subseteq \dots$$

is an ideal of R .

Proof: Let $A = \bigcup_i A_i$. As $A_i \subseteq A \forall i$, A is non-empty. Let $a, b \in A$, $r \in R$. Then $a \in A_i$, $b \in A_j$ for some i and j . Now either $i \leq j$ or $j < i$. To be definite let $i \leq j$. Since $A_1 \subseteq A_2 \subseteq A_3 \subseteq \dots$ holds we have $A_i \subseteq A_j$. Consequently $a, b \in A_j$. This yields $a - b \in A_j$, $ar \in A_j$, $ra \in A_j$ as A_j is an ideal of R . Hence $a - b \in A$, $ar \in A$, $ra \in A$ as $A_i \subseteq A$. Hence

A is an ideal of R . ■

Lemma 10.27. *In a PID, for every ascending chain of ideals*

$$A_1 \subseteq A_2 \subseteq A_3 \subseteq \dots \subseteq A_n \subseteq \dots$$

there exists an integer t such that $A_m = A_t \forall m \geq t$.

Proof: Let $A = \cup_i A_i$. A is an ideal of R by Lemma 10.26. As R is a PID, $A = \langle a \rangle$ for some $a \in A$. Now $a \in A_t$ for some t . Consequently $A = \langle a \rangle \subseteq A_t$. Then for every integer $m \geq t$, $A \subseteq A_t \subseteq A_m \subseteq A$ yields $A_m = A_t$. Hence the lemma follows. ■

Lemma 10.28. *For every non-zero non-unit element a in a PID there exists an irreducible element p such that $p \mid a$.*

Proof: Let a be a non-zero and non-unit element of a ring R which is a PID. Consider the ideal $I_1 = \langle a \rangle$. If I_1 is maximal then a is irreducible (Theorem 10.13) and there is nothing to prove. If I_1 is not maximal, there exists an ideal I_2 of R such that $I_1 < I_2 < R$. If I_2 is maximal then it is generated by an irreducible element, say p_1 . This gives $I_1 = \langle a \rangle < I_2 = \langle p_1 \rangle \Rightarrow p_1 \mid a$ and we are done. Otherwise we can find an ideal I_3 of R such that $I_1 < I_2 < I_3 < R$. This process cannot continue indefinitely in view of Lemma 10.27, so there exists a maximal ideal I_n such that $I_1 < I_2 < I_3 < \dots < I_n < R$. Then I_n is generated by an irreducible element, say, p . This implies $I_1 = \langle a \rangle < I_n = \langle p \rangle \Rightarrow p \mid a$. Hence the lemma follows. ■

Theorem 10.29. *Every PID is UFD.*

Proof: Let a be a non-zero non-unit element of a ring R which is a PID. There is an irreducible element p_1 (which is also prime by Theorem 10.12) such that $p_1 \mid a$ i.e. $a = a_1 p_1$ for some $a_1 \in R$. So $\langle a \rangle \subseteq \langle a_1 \rangle$. If $\langle a_1 \rangle = \langle a \rangle$ then $a_1 = ar$ for some $r \in R \Rightarrow a = ar p_1 \Rightarrow r p_1 = 1 \Rightarrow p_1$ is unit, which is absurd as p_1 is an irreducible element. So $\langle a \rangle < \langle a_1 \rangle$. Now if a_1 is unit then a is an associate of a irreducible element and hence it is irreducible itself and there is nothing to prove. If a_1 is not a unit then there exists an irreducible element p_2 such that $p_2 \mid a_1$ i.e. $a_1 = a_2 p_2$ for some $a_2 \in R$. This gives $a = a_2 p_1 p_2$ and $\langle a_1 \rangle < \langle a_2 \rangle$.

If a_2 is a unit we stop, otherwise we can write as before $a_2 = a_3 p_3$ for some irreducible element p_3 and some non-unit a_3 of R . Here again $\langle a_2 \rangle < \langle a_3 \rangle$. In this way we get a properly (or strictly) ascending chain of ideals

$$\langle a \rangle < \langle a_1 \rangle < \langle a_2 \rangle < \langle a_3 \rangle < \dots$$

Such a chain of ideals in a PID must be finite (Lemma 10.27). Thus for some n , a_{n+1} must be unit and as $a_n = a_{n+1} q_{n+1}$ for some irreducible element q_{n+1} of R , we must have a_n itself irreducible. Then $a = a_1 p_1, a_1 = a_2 p_2, \dots, a_{n-1} = p_n a_n = p_n p_{n+1}$ where $a_n = p_{n+1}$, this gives

$$a = p_1 p_2 \dots p_{n+1} \dots \quad (1)$$

a product of finitely many irreducible elements. Thus every non-unit non-zero element of R is expressible as a product of finite number of irreducible elements.

Let $a = q_1 q_2 \dots q_m$... (2)
 be another expression of a as a product of a finite number of irreducible elements. We want to show that (1) and (2) are equivalent expressions. For this we apply induction on $n+1$. Note that for $n=0$ the result holds trivially. Suppose that the result holds for all those non-zero non-units which are expressible as products of less than $(n+1)$ irreducible elements. Now (2) $\Rightarrow q_1 \mid a \Rightarrow q_1 \mid p_1 p_2 \dots p_{n+1}$. Since every irreducible element in a PID is prime, we get $q_1 \mid p_i$ for some i . As p_i is irreducible, we have $q_1 = u_i p_i$ for some unit u_i . By re-indexing the p_i 's we can suppose that $i=1$. Then from (1) and (2), it follows that $p_1 p_2 p_3 \dots p_{n+1} = u_1 p_1 q_2 \dots q_m \Rightarrow p_2 p_3 \dots p_{n+1} = q_2' q_3' \dots q_m'$ where $q_2' = u_1 q_2$ and $q_i' = q_i \forall i \geq 3$. If we put $b = p_2 p_3 \dots p_{n+1}$, we see that

$$b = p_2 p_3 \dots p_{n+1} \dots (3)$$

$$\text{and } b = q_2' q_3' \dots q_m' \dots (4)$$

are two expressions of b as products of irreducible elements and (3) has n terms. Since $n < n+1$, the induction hypothesis gives $n = m-1$ i.e. $n+1 = m$ and there exists a 1-1 correspondence between p_i 's ($i=2, 3, \dots, n+1$) and q_j' 's ($j=2, 3, \dots, n+1$) such that the corresponding elements are associates. Since, as seen before $q_1 = u_1 p_1$ with u_1 as a unit, q_1 is also an associate of p_1 . Hence in (1) and (2) the number of factors are equal and there exist a 1-1 correspondence between p_i 's and q_j 's such that the corresponding elements are associates. This proves the theorem. ■

Since every Euclidean Domain is a PID, we get

Corollary 10.30. Every Euclidean Domain is a UFD. ■

As for any field F , the polynomial ring $F[x]$, and the ring of Gaussian integers are all PID's we obtain from above theorem that all these are also UFD's.

Lemma 10.31. In a UFD every pair of non-zero elements have an HCF and LCM.

Proof: Let R be a UFD and let $a, b \in R$ be two non-zero elements.

Consider the case when one of a and b , say a , is a unit. Then a^{-1} exists and $b = aa^{-1}b \Rightarrow a \mid b$. However $a \mid a$, so a is a common factor of a and b . Let $d \in R$ such that $d \mid a$ and $d \mid b$. Then $d \mid a$ and so a is HCF of a and b . Again $a \mid b$ and $b \mid b$ gives b is a common multiple of a and b . Let $c (\neq 0) \in R$ be such that $a \mid c$ and $b \mid c$. This gives $b \mid c$. Hence b is LCM of a and b . Similarly if b is a unit then b is an HCF of a and b , and a is an LCM of a and b .

So we suppose that a and b both are non-units. In this case we can express

$$a = p_1^{\alpha_1} p_2^{\alpha_2} \dots p_k^{\alpha_k} \quad \dots (1)$$

$$b = p_1^{\beta_1} p_2^{\beta_2} \dots p_k^{\beta_k} \quad \dots (2)$$

as products of powers of same set of distinct primes. (Here we allow zero powers as well, e.g. $2 = 2^1 \cdot 3^0$, $3 = 2^0 \cdot 3^1$ in $\mathbb{Z}$). For every $i = 1, 2, 3, \dots, k$ let $\gamma_i = \max(\alpha_i, \beta_i)$ and $\delta_i = \min(\alpha_i, \beta_i)$. Further let $c = p_1^{\delta_1} p_2^{\delta_2} \dots p_k^{\delta_k}$ and $d = p_1^{\gamma_1} p_2^{\gamma_2} \dots p_k^{\gamma_k}$. $\dots (3)$

Now notice that given any element $x = p_1^{\lambda_1} p_2^{\lambda_2} \dots p_k^{\lambda_k}$ in R , if $y \mid x$ then we can express $y = up_1^{\mu_1} p_2^{\mu_2} \dots p_k^{\mu_k}$ for non-negative integers $\mu_i \leq \lambda_i$ and some unit $u \in R$ and conversely (Why?). By definition of δ_i , $\delta_i \leq \alpha_i \forall i$ and hence $c \mid a$. Similarly $c \mid b$. If $c' \in R$ is such that $c' \mid a$ and $c' \mid b$ then we must have $c' = v p_1^{\theta_1} p_2^{\theta_2} \dots p_k^{\theta_k}$ for some $\theta_i \leq \alpha_i, \beta_i \forall i$ and some unit v . $\dots (4)$

Thus $\theta_i \leq \min(\alpha_i, \beta_i) = \delta_i \forall i$. Further as v is a unit, we get from (3) and (4), $c' \mid c$. Hence c is an HCF of a and b .

Similarly one can prove that d is an LCM of a and b . ■

An obvious application of induction yields.

Lemma 10.38. Any finite number of non-zero elements $a_1, a_2, \dots, a_n$ of a UFD have an HCF and an LCM. ■

We now study the ring of polynomials over a UFD and prove a number of results in the forms of lemmas and theorems which are in themselves of much interest.

Definition 10.33. Let R be UFD and let $f = a_0 + a_1x + a_2x^2 + \dots + a_nx^n$ be non-zero polynomial in $R[x]$. Let d be an HCF of the coefficients $a_0, a_1, a_2, \dots, a_n$ of f . Then the ideal dR is called the *content* of f and we denote the same by $C(f)$.

Notice that $C(f)$ is uniquely defined. For if d' is another HCF of $a_0, a_1, a_2, \dots, a_n$ then d' is an associate of d and so $d' = du$ for some unit $u \in R$. This gives $d'R = duR = dR$ since $uR = R$. Hence $C(f)$ is uniquely defined. Note further that if an element a divides every coefficient of f then $a \mid d$ where d is an HCF of the coefficients of f .

REMARK. Some authors define any HCF of the coefficients of a non-zero polynomial as its content. Since HCF is not unique, this definition does not give unique content. In fact in recent works, the above definition of content in finding popularity.

Definition 10.34. Let R be a UFD then a non-zero polynomial $f \in R[x]$ is said to be a *primitive polynomial* if HCF of its coefficients is a unit.

Now given $d \in R$, $dR = R$ if and only if d is a unit; so we can say that a polynomial $f \in R[x]$ is primitive if and only if $C(f) = R$.

EXAMPLE 22. For $f = 2 + 6x + 10x^2 + 18x^3 \in \mathbb{Z}[x]$, $C(f) = E$, the set of all even integers. As $E \neq \mathbb{Z}$, f is not primitive. (Note that HCF of 2, 6, 10 and 18 is 2 which is not a unit in $\mathbb{Z}$).

EXAMPLE 23. In $\mathbb{Z}[x]$ the polynomial $g = 3 + 5x + 4x^3 + 15x^4$ is primitive since HCF of 3, 5, 4 and 15 is 1.

Lemma 10.35. If R is a UFD then every non-zero polynomial in $R[x]$ is a product of a primitive polynomial over R and an element of R .

Proof: Let $f = a_0 + a_1x + a_2x^2 + \dots + a_nx^n$ be a non zero polynomial in $R[x]$, where R is a UFD. Let d be an HCF of $a_0, a_1, a_2, \dots, a_n$ in R . Then $a_i = db_i$ for some $b_i \in R \forall i = 0, 1, 2, \dots, n$. Since d is an HCF of $a_0, a_1, \dots, a_n$, an HCF of $b_0, b_1, \dots, b_n$ is 1. This proves that $f = d(b_0 + b_1x + b_2x^2 + \dots + b_nx^n) = dg$ with $g = b_0 + b_1x + \dots + b_nx^n$ and HCF of $b_0, b_1, b_2, \dots, b_n$ is 1. Hence g is primitive. Thus the lemma is proved. ■

Theorem 10.36. (Gauss). The product of two primitive polynomials over a UFD is a primitive polynomial.

Proof: Let R be a UFD and let

$$f = a_0 + a_1x + a_2x^2 + \dots + a_nx^n \quad \dots (1)$$

$$g = b_0 + b_1x + b_2x^2 + \dots + b_mx^m \quad \dots (2)$$

be two primitive polynomials over R . Then

$$fg = c_0 + c_1x + c_2x^2 + \dots + c_{m+n}x^{m+n} \quad \dots (3)$$

where

$$c_k = \sum_{i+j=k} a_ib_j \text{ for } 0 \leq k \leq m+n.$$

Let $d = (c_0, c_1, c_2, \dots, c_{m+n})$ be a non-unit, then there exists an irreducible element p such that $p \mid d$. Clearly $p \mid c_k \forall k$. Since HCF of $a_0, a_1, a_2, \dots, a_n$ is a unit (as f is primitive) p cannot divide all a_i 's. Let t be the smallest suffix such that $p \nmid a_t$. Similarly let u be the smallest suffix such that $p \nmid b_u$.

$$\text{Consider } c_{t+u} = (a_0b_{t+u} + a_1b_{t+u-1} + \dots + a_{t-1}b_{u+1}) + a_tb_u + (a_{t+1}b_{u-1} + a_{t+2}b_{u-2} + \dots + a_{t-1}b_u) \quad \dots (4)$$

By our choice $p \mid a_0, p \mid a_1, p \mid a_2, \dots, p \mid a_{t-1}$. This yields $p \mid (a_0b_{t+u} + a_1b_{t+u-1} + \dots + a_{t-1}b_{u+1})$. Similarly $p \mid (a_{t+1}b_{u-1} + a_{t+2}b_{u-2} + \dots + a_{t-1}b_u)$. Since also $p \mid c_{t+u}$, (4) gives $p \mid a_tb_u \Rightarrow p \mid a_t$ or $p \mid b_u$ as p is prime (Lemma 10.25). This is a contradiction. Hence fg is primitive. ■

Corollary 10.37. For two non-zero polynomials f and g over a UFD, $C(fg) = C(f)C(g)$.

Proof: Let R be a UFD and $f, g \in R[x]$ be two non-zero polynomials; let d and d' be HCF 's of the coefficients of f and g respectively. Then $f = df_1$ and $g = d'g_1$,
 $\dots (1)$

where f_1 and g_1 are primitive polynomials, $C(f)=dR$ and $C(g)=d'R$. Then $fg=(dd')f_1g_1$ (2)

Now as f_1g_1 is primitive (Theorem 10.36) dd' is an HCF of the coefficients of fg . Consequently $C(fg)=dd'R=(dR)(d'R)=C(f)C(g)$ as R is commutative. ■

REMARK. The above corollary immediately gives that if $f=gh$ and one of g and h is not primitive then f is not primitive.

Definition 10.38. Let R be an integral domain with unity. A polynomial f over R of positive degree is said to be *irreducible polynomial* over R if f cannot be expressed as a product of two other polynomials over R both of positive degree. Any polynomial of positive degree which is not an irreducible polynomial over R is said to be *reducible* over R .

Clearly any polynomial of degree 1 over any integral domain R is irreducible polynomial over R .

EXAMPLE 24. x^2+1 , x^2-2 , x^2+x+1 are all irreducible polynomials in $\mathbf{Z}[x]$.

EXAMPLE 25. x^2+3 is an irreducible polynomial in $\mathbf{R}[x]$.

EXAMPLE 26. x^2+1 is reducible polynomial in $\mathbf{C}[x]$, as $x^2+1=(x+i)(x-i)$, since $i=\sqrt{-1} \in \mathbf{C}$.

Lemma 10.39. For any integral domain R with unity, the units of $R[x]$ are precisely the units of R .

Proof: Clearly a unit of R is a unit of $R[x]$ as $R \subseteq R[x]$ and they have same unity.

Conversely let $f \in R[x]$ be a unit in $R[x]$ then there exists $g \in R[x]$ such that $fg=1$. This implies $\deg f + \deg g = 0 \Rightarrow \deg f = \deg g = 0 \Rightarrow f$ and g are constants. Thus f is a unit of R . Hence the lemma follows. ■

Lemma 10.40. If R is a UFD then any $f \in R[x]$ is an irreducible element of $R[x]$ if and only if either f is an irreducible element of R or f is an irreducible primitive polynomial in $R[x]$.

Note. We advise the readers to keep in mind the difference between irreducible element of $R[x]$ and irreducible polynomial in $R[x]$.

Proof: Let $f \in R[x]$ be an irreducible element. If $f \in R$ then trivially f is irreducible element of R ; further if $f \notin R$ let $f=gh$ for some $g, h \in R[x]$. As f is an irreducible element of $R[x]$, one of g and h must be a unit of $R[x]$, so one of g and h must be in R (Lemma 10.39). Consequently $\deg g=0$ or $\deg h=0$. This means that f is an irreducible polynomial in $R[x]$. As $f=df_1$ where $d \in R$ and f_1 is primitive polynomial (Lemma 10.35) and $\deg f_1 = \deg f > 0$, f_1 is not a unit, so d must be a unit. Hence f is primitive.

Conversely any irreducible element of R is an irreducible element

of $R[x]$. Let f be an irreducible primitive polynomial in $R[x]$ and let $f=gh$ for some $g, h \in R[x]$. Since f is an irreducible polynomial either $\deg g=0$ or $\deg h=0$. To be definite let $\deg g=0$.

Then $g=\alpha(\neq 0) \in R$ and so $f=\alpha h$. Since $C(f)=1R$ we get $\alpha \mid 1$, hence α is a unit of R so of $R[x]$. This gives f is an irreducible element of $R[x]$. ■

REMARK 1. Let K be the field of quotients of an integral domain R . Consider any element y in an overring of K . Suppose y is transcendental over R . Consider a non-zero polynomial $f=a_0+a_1y+\dots$

$+a_ny^n$ ($a_n \neq 0$) over K . Now for all $i=0, 1, 2, \dots, n$; $a_i = \frac{c_i}{d_i}$ for some $c_i, d_i \in R$ with $d_i \neq 0$. Put $d=d_0d_1d_2\dots d_n$. Then

$$df=da_0+da_1y+\dots+da_iy^i+\dots+da_ny^n.$$

Now for any $i=0, 1, 2, \dots, n$; $da_i=d_0d_1\dots d_n \cdot \frac{c_i}{d_i}=d_0d_1d_2\dots d_{i-1}d_{i+1}\dots d_n c_i \in R$. Thus $df(y) \in R[y]$.

Now if $f(y)=0$ then $df(y)=0$. Since y is transcendental over R $df(y) \neq 0$ and so $f(y) \neq 0$. Hence y is transcendental over K .

REMARK 2. In case R is a field, as every non zero element of R is a unit, it cannot be an irreducible element, the above lemma gives that irreducible elements in $R[x]$ are precisely the irreducible polynomials in $R[x]$.

Lemma 10.41. Let R be a UFD and K be its field of quotients then an irreducible primitive polynomial in $R[x]$ is also an irreducible polynomial in $K[x]$.

Proof: Let f be an irreducible primitive polynomial in $R[x]$ where R is a UFD. Let $f=gh$ for some g and h in $K[x]$ with $\deg g > 0$ and $\deg h > 0$. Since K is the field of quotients of R , there exist non-zero elements d, d' in R such that dg and $d'h$ both belong to $R[x]$ (see remark 1). This gives $dd'f=(dg)(d'h)$.

Now $dg=\alpha g_1$ and $d'h=\beta h_1$ where g_1 and h_1 are some primitive polynomials in $R[x]$ and $\alpha, \beta \in R$.

Then $dd'f=(\alpha\beta)g_1h_1$ with f and g_1h_1 primitive in $R[x]$. This yields $udd'=\alpha\beta$ for some unit u in R . This implies $f=ug_1h_1$. However f is irreducible polynomial in $R[x]$. Therefore, either $\deg ug_1=0$ or $\deg h_1=0$. However $\deg ug_1=\deg g > 0$ and $\deg h_1=\deg h > 0$, so we arrive at a contradiction. ■

REMARK. In view of Remark 2 preceding the above lemma, we can say that every irreducible primitive polynomial in $R[x]$ is an irreducible element of $K[x]$ where K is the field of quotients of R and R is a UFD.

Lemma 10.42. If R is a UFD and K is its field of quotients then

(1) every irreducible element of R is a prime element of $R[x]$ and (2) if f is a primitive polynomial in $R[x]$ and $h \in R[x]$ is such that $f \mid h$ in $K[x]$ then $f \mid h$ in $R[x]$.

Proof: (1) Let p be an irreducible element of R and let f, g be two non-zero polynomials in $R[x]$ such that $p \mid fg$ in $R[x]$. Then $fg = ph$ for some $h \in R[x]$. Now let $C(f) = dR$, $C(g) = d'R$ and $C(h) = lR$ for some $d, d', l \in R$. Thus $f = df_1$, $g = d'g_1$ and $h = lh_1$ with f_1, g_1 and h_1 primitive polynomials in $R[x]$. Then $fg = ph \Rightarrow dd'f_1g_1 = plh_1 \Rightarrow dd'R = plR$ as $C(fg) = dd'$ and $C(ph) = plR$. This in turn implies $dd' = plu$ for some unit $u \in R \Rightarrow p \mid dd'$ in R . As every irreducible element in R is prime we obtain $p \mid d$ or $p \mid d'$ in R and so the same holds in $R[x]$. However $d \mid f$ and $d' \mid g$ in $R[x]$. Hence $p \mid f$ or $p \mid g$ in $R[x]$. This proves that p is a prime element of $R[x]$.

(2) Since $f \mid h$ in $K[x]$ there exists $g \in K[x]$ such that $h = fg$. As K is the field of quotients of R there exists $d (\neq 0) \in R$ such that $dg \in R[x]$. Then $dh = dfg$. Now we can write $h = \alpha h_1$ and $dg = \beta g_1$ for $\alpha, \beta \in R$ and primitive polynomials h_1 and g_1 in $R[x]$. This gives $d\alpha h_1 = f\beta g_1$. As h_1 and fg_1 are primitive polynomials in $R[x]$, $d\alpha$ and β are associates in $R \Rightarrow \beta = d\alpha u$ for some unit $u \in R$. This in turn gives $h_1 = ufg_1$. Hence $f \mid h_1$ in $R[x]$. But $h_1 \mid h$ in $R[x]$.

Consequently $f \mid h$ in $R[x]$. ■

Corollary 10.43. If R is a UFD then every irreducible element of $R[x]$ is a prime element of $R[x]$.

Proof: Let $f \in R[x]$ be an irreducible element. Either f is irreducible element of R or f is an irreducible primitive polynomial in $R[x]$ (Lemma 10.40). In the first case f is a prime element of $R[x]$, [Lemma 10.42 (1)]. In the second case f is irreducible polynomial in $K[x]$ where K is the field of quotients of R (Lemma 10.41). As K is a field, f is an irreducible element of $K[x]$. However $K[x]$ is a PID (Theorem 10.11). So f is prime element of $K[x]$. Now let $f \mid gh$ in $R[x]$ then $f \mid gh$ in $K[x]$. But f being a prime element of $K[x] \Rightarrow$ either $f \mid g$ in $K[x]$ or $f \mid h$ in $K[x]$. Using Lemma 10.42 (2), we get either $f \mid g$ or $f \mid h$ in $R[x]$. Hence f is a prime element of $R[x]$. ■

We are now fully equipped to attack our target; the celebrated theorem on UFD.

Theorem 10.44. If R is a UFD then $R[x]$ is a UFD.

Proof: Clearly $R[x]$ is an integral domain with unity. Firstly, we show that every non-zero non-unit element of $R[x]$ can be written as a product of finite number of irreducible elements of $R[x]$. Let f be non-zero non-unit element of $R[x]$. If $f \in R$ then we are through, as R is a UFD and irreducible elements of R are irreducible elements of $R[x]$. To discuss the case $f \notin R$ firstly we prove our assertion for primitive

polynomials f in $R[x]$ of positive degree, by induction on $\deg f$. If $\deg f$ is 1 then f is an irreducible primitive polynomial and so is an irreducible element of $R[x]$. Let $\deg f = n > 1$ and let our assertion be true for all primitive polynomials of degree $< n$. Again if f is irreducible we are done, so let f be reducible. Then $f = gh$ for some $g, h \in R[x]$ with $\deg g < n$ and $\deg h < n$. Also as f is primitive polynomial, g and h are primitive polynomials of degrees properly less than n , so by induction hypothesis both g and h are expressible as a product of finite number of irreducible elements of $R[x]$; as a consequence $f = gh$ is also expressible as a product of finite number of irreducible elements of $R[x]$.

In case f is not primitive, we can write $f = df_1$ where f_1 is a primitive polynomial in $R[x]$ and d is a non-unit in R . As seen above both d and f_1 can be expressed as products of finite number of irreducible elements of $R[x]$. Hence $R[x]$ satisfies the first condition for an integral domain with unity, to be a UFD.

Now let $f = f_1 f_2 \dots f_m$ and $f = g_1 g_2 \dots g_n$ be two expressions of f as products of finitely many irreducible elements of $R[x]$. We apply induction on m . For $m = 1$, $f = f_1 = g_1 g_2 \dots g_n$. Since $g_1 \mid f_1$, so if $n > 1$, we get $tg_2 \dots g_n = 1$ for some $t \in R$. This implies g_n is a unit which is absurd. Thus $n = 1$ and $f_1 = g_1$. Suppose $m > 1$ and let any two decompositions be equivalent for all those non-zero non-unit elements in $R[x]$ which are expressible as products of less than m irreducible elements of $R[x]$.

Now $g_1 \mid f \Rightarrow g_1 \mid f_1 f_2 \dots f_m$. As g_1 is a prime element of $R[x]$ (Corollary 10.43); we get $g_1 \mid f_i$ for some i .

Since f_i is also an irreducible element we have $g_1 = u_i f_i$ for some unit u_i in $R[x]$ which must be in R . By re-labelling the f_i 's, we can suppose that $i = 1$.

Then it follows that $f_1 f_2 f_3 \dots f_m = u_1 f_1 g_2 g_3 \dots g_n \Rightarrow f_2 f_3 \dots f_m = g_2' g_3' \dots g_n'$ where $g_2' = u_1 g_2$ and $g_k' = g_k \forall k \geq 3$. If we put $h = f_2 f_3 \dots f_m$ then we see that $h = f_2 f_3 \dots f_m$ and $h = g_2' g_3' \dots g_n'$ are two expressions of h as products of irreducible elements and the first decomposition has $m - 1$ irreducible elements. So by the induction hypothesis these two decompositions are equivalent, i.e. $n - 1 = m - 1$ (so $n = m$) and there exists a 1-1 correspondence between f_i 's ($i = 2, 3, \dots, m$) and g_j 's ($j = 2, 3, \dots, m$) such that the corresponding elements are associates.

Again as $g_1 = u_1 f_1$ with u_1 , a unit, g_1 is also an associate of f_1 . Hence there is a 1-1 correspondence between f_i 's ($i = 1, 2, \dots, m$) and g_j 's ($j = 1, 2, \dots, m$) such that the corresponding elements are associates. This shows that $R[x]$ satisfies condition (2) for a UFD. Consequently $R[x]$ is a UFD. ■

EXAMPLE 27. As $\mathbf{Z}$ is a UFD, $\mathbf{Z}[x]$ is a UFD (Theorem 10.44). We claim that $\mathbf{Z}[x]$ is not a PID. For this purpose consider the ideal $I = \langle x \rangle + \langle 4 \rangle$. If I is a principal ideal, we can find a polynomial $f \in \mathbf{Z}[x]$ such that $I = \langle f \rangle$ i.e. $\langle x \rangle + \langle 4 \rangle = \langle f \rangle$. This gives $\langle 4 \rangle \subseteq \langle f \rangle$ so that $f \mid 4$ in $\mathbf{Z}[x]$ i.e. $4 = fg$ for some: $g \in \mathbf{Z}[x]$. However $\deg 4 = 0 \Rightarrow \deg f = 0 \Rightarrow f \in \mathbf{Z}$. Let $f = \alpha (\neq 0) \in \mathbf{Z}$; then $\langle x \rangle + \langle 4 \rangle = \langle \alpha \rangle$. Further $\langle x \rangle \subseteq \langle \alpha \rangle \Rightarrow x = \alpha h$ for some: $h \in \mathbf{Z}[x] \Rightarrow \deg h = 1 \Rightarrow h = \beta x + \gamma$ for some $\beta, \gamma \in \mathbf{Z}$ with $\beta \neq 0$. Now: $x = \alpha h \Rightarrow \alpha \beta = 1$. Thus $1 \in \langle \alpha \rangle \Rightarrow 1 \in \langle x \rangle + \langle 4 \rangle \Rightarrow 1 = xt + 4v$ for some $t, v \in \mathbf{Z}[x]$. But this in turn implies that 4 divides 1 in $\mathbf{Z}$ (Put $v = \beta_0 + \beta_1 x + \dots + \beta_m x^m$, $\beta_i \in \mathbf{Z}$ in $1 = xt + 4v$ and compare coefficients), which is impossible hence I is not a principal ideal. Consequently $\mathbf{Z}[x]$ is not a PID.

Now at the end of the chapter we state and prove a test for the irreducibility of a polynomial in $\mathbf{Z}[x]$ over $\mathbf{Q}$.

Theorem 10.45. (Eisenstein's Irreducibility Criterion). Let $f = a_0 + a_1 x + \dots + a_n x^n$ be a polynomial in $\mathbf{Z}[x]$. If there exists a prime number p such that $p \mid a_i$ for all $i = 0, 1, 2, \dots, n-1$ and $p \nmid a_n$, and $p^2 \nmid a_0$, then f is an irreducible polynomial over $\mathbf{Q}$.

Proof: Note that the given conditions imply that $\deg f > 0$. We firstly show that f is an irreducible polynomial in $\mathbf{Z}[x]$. If not, let $f = gh$ with $g, h \in \mathbf{Z}[x]$ such that $\deg g < \deg f$ and $\deg h < \deg f$.

Let $g = b_0 + b_1 x + b_2 x^2 + \dots + b_s x^s$ with $b_s \neq 0$
and $h = c_0 + c_1 x + c_2 x^2 + \dots + c_t x^t$ with $c_t \neq 0$

Now $f = gh \Rightarrow a_0 = b_0 c_0 \Rightarrow p \mid b_0$ or $p \mid c_0$, as it is given that $p \mid a_0$. However as $p^2 \nmid a_0$, p cannot divide both b_0 and c_0 . To be definite let $p \mid b_0$ and $p \nmid c_0$. Again $f = gh \Rightarrow a_n = b_s c_t$. As $p \nmid a_n$ we have $p \nmid b_s$. So let b_k be the first coefficient of g (from left) which is not divisible by p . Clearly $k > 0$ and $k \leq s < n$. Now $a_k = b_k c_0 + b_{k-1} c_1 + \dots + \dots + b_0 c_k$, is such that $p \mid a_k$, $p \mid b_0$, $p \mid b_1$, $p \mid b_2, \dots, p \mid b_{k-1}$. Consequently $p \mid b_k c_0$. As $p \nmid c_0$ and p is prime, $p \mid b_k$. This is a contradiction, hence f is an irreducible polynomial in $\mathbf{Z}[x]$. Now we can write $f = d f_1$ where $d \in \mathbf{Z}$ and f_1 is a primitive polynomial in $\mathbf{Z}[x]$. Then f_1 is also irreducible in $\mathbf{Z}[x]$. This gives f_1 is an irreducible in $\mathbf{Q}[x]$ as $\mathbf{Z}$ is a UFD and $\mathbf{Q}$ is its field of quotients (Lemma 10.41). But $f_1 = \left(\frac{1}{d}\right) f \in \mathbf{Q}[x]$ so f is also irreducible in $\mathbf{Q}[x]$. ■

EXAMPLE 28. The polynomials $x^4 - 4x + 2$, $x^3 - 9x + 15$, $7x^4 - 2x^3 + 6x^2 - 10x + 18$ are all irreducible polynomials in $\mathbf{Q}[x]$. In the case of $x^4 - 4x + 2$, $p = 2$ satisfies Einstein's Criterion. In the case of $x^3 - 9x + 15$, $p = 3$ and finally in the case of $7x^4 - 2x^3 + 6x^2 - 10x + 18$, $p = 2$

satisfy Einstein's Criterion.

EXAMPLE 29. Consider $f(x) = x^2 + 1$. Then $f(x+1) = x^2 + 2x + 2$ taking $p=2$ we see that Eisenstein's Criterion is applicable to $f(x+1)$. Since $f(x) = g(x)h(x)$ if and only if $f(x+a) = g(x+a)h(x+a) \forall a \in \mathbb{Z}$ we get $x^2 + 1$ is irreducible over $\mathbb{Q}$. Note that there is no prime p such that $p \mid 1$. This shows that Eisenstein's Criterion is *not necessary* for the irreducibility of any polynomial in $\mathbb{Z}[x]$ over $\mathbb{Q}$.

Notice that in general if R is a UFD and $f \in R[x]$ then for any $a \in R$, $f[x]$ is irreducible polynomial over R if and only if $f(x+a)$ is an irreducible polynomial over R . Since $f(x) = h(x)g(x) \Rightarrow f(x+a) = g(x+a)h(x+a)$ and $\deg g(x) = \deg g(x+a)$, $\deg h(x) = \deg h(x+a)$. So sometimes Eisenstein's Criterion can be applied successfully if we replace x in a polynomial over $\mathbb{Z}$ by $x+n$ where n is suitable integer (usually we take $n = \pm 1, \pm 2$).

EXAMPLE 30. For any prime number p , the polynomial $x^{p-1} + x^{p-2} + \dots + x^2 + x + 1$ is irreducible over $\mathbb{Q}$.

Let $f(x) = x^{p-1} + x^{p-2} + \dots + x^2 + x + 1 = \frac{x^p - 1}{x - 1}$. Then

$$f(x+1) = \frac{(x+1)^p - 1}{(x+1) - 1} = \frac{x^p + {}^pC_1 x^{p-1} + \dots + {}^pC_r x^{p-r} + \dots + {}^pC_{p-1} x}{x} \\ = x^{p-1} + {}^pC_1 x^{p-2} + \dots + {}^pC_r x^{p-r-1} + \dots + {}^pC_{p-1}.$$

Now as p is a prime number, $p \nmid {}^pC_r \forall 1 \leq r \leq p-1$. (Prove!). Also ${}^pC_{p-1} = p$; so $p^2 \nmid {}^pC_{p-1}$.

Hence by Eisenstein's Criterion, $f(x+1)$ is an irreducible polynomial in $\mathbb{Q}[x]$ and so is $f(x)$ in $\mathbb{Q}[x]$.

PROBLEMS

1. In $\mathbb{Z}[\sqrt{-5}] = \{a + b\sqrt{-5} \mid a, b \in \mathbb{Z}\}$ show that $3, 1 + 2\sqrt{-5}$ are co-prime but have no LCM. Also show that in $\mathbb{Z}[\sqrt{-5}]$ the elements $21(1 - 2\sqrt{-5})$ and $7(1 + 2\sqrt{-5})(1 - 2\sqrt{-5})$ have no HCF.

2. Show that the set of polynomials in $\mathbb{Z}[x]$ with even coefficients of x , form a subring R . Verify that 2 and $2x$ have an HCF but no LCM in R .

3. Show that $\mathbb{Z}[\sqrt{-6}] = \{a + b\sqrt{-6} \mid a, b \in \mathbb{Z}\}$ is not a UFD.

[Hint. $10 = 2 \cdot 5 = (2 + \sqrt{-6})(2 - \sqrt{-6})$.]

4. Show that $\mathbb{Z}[\sqrt{-7}]$ is not a UFD.

[Hint. $16 = 2 \cdot 2 \cdot 2$ and also $16 = (3 + \sqrt{-7})(3 - \sqrt{-7})$.]

5. Let R be a UFD and f , a polynomial of positive degree in $R[x]$, prove that $f = d_1 f_1$ for some $d_1 \in R$ and f_1 , a primitive polynomial in $R[x]$. If $f = d_1 f_1 = d_2 f_2$ with $d_1, d_2 \in R, f_1, f_2$ two primitive poly-

nomials then prove that d_1 and d_2 are associates in R and f_1, f_2 are associates in $R[x]$.

6. Prove that if p is a prime number then for all $1 \leq r \leq p-1$, p divides ${}^pC_r$.

7. If R is a UFD and if K is the field of quotients of R . Prove that two primitive polynomials f and $g \in R[x]$ are associates in $R[x]$ if and only if they are associates in $K[x]$.

8. Let a, b and c be three non-zero elements of a ring R which is a UFD. Prove the following:

$$(a, [b, c]) = [(a, b), (a, c)]$$

and

$$[a, (b, c)] = ([a, b], [a, c])$$

In other words HCF operation is distributive over LCM operation and vice versa.

9. Prove that in a UFD all minimal prime ideals are principal and are exactly those ideals which are generated by irreducible elements.

10. Show that if p is a prime number, then the polynomial.

$$x^{p(p-1)} + x^{p(p-2)} + \dots + x^p + 1$$

is irreducible over $\mathbb{Q}$.

$$\left[\text{Hint. The given polynomial} = \frac{x^{p^2} - 1}{x^p - 1} \right]$$

11. A, B, C are ideals of a PID, R , Show that

$$(a) \quad A \cap (B + C) = A \cap B + A \cap C$$

$$(b) \quad A + (B \cap C) = (A + B) \cap (A + C)$$

12. R is a UFD and P is a prime ideal of R such that P is generated by a non-unit x . Show that there is no prime ideal Q of R such that $(0) \subsetneq Q \subsetneq R$.

[Hint. First show that if such a Q exists then $Q = Ax$ for some ideal A of R . Next show that we can find an irreducible element p such that $p^n \mid x$ for all integers $n \geq 1$.]

13. R is a UFD and a, b, c are non-zero elements of R . Prove

(i) If $a \mid c$ and $b \mid c$ and $(a, b) = 1$ then $ab \mid c$.

(ii) If $a \mid bc$ and $(a, b) = 1$ then $a \mid c$.

MISCELLANEOUS WORKED-OUT EXERCISES

Exercise 1. Prove that $x^2 + 1$ and $x^2 + x + 4$ are irreducible over F , the field of integers modulo 11. Prove further that $\frac{F[x]}{\langle x^2 + 1 \rangle}$ and

$\frac{F[x]}{\langle x^2+x+4 \rangle}$ are isomorphic fields, each having 121 elements.

Solution. The elements of F are $\bar{0}, \bar{1}, \bar{2}, \bar{3}, \bar{4}, \bar{5}, \bar{6}, \bar{7}, \bar{8}, \bar{9}$ and $\bar{10}$. For convenience sake we shall drop the 'bar' sign. None of the polynomials x^2+1 and x^2+x+4 are satisfied by the elements of F . As a result of this none of them has any linear factor in $F[x]$. This shows the irreducibility of both the polynomials over F . Any element of

$\frac{F[x]}{\langle x^2+1 \rangle}$ is of the type $f(x) + \langle x^2+1 \rangle$, with $f(x) \in F[x]$. By Euclidean algorithm in $F[x]$, $f(x) = g(x)(x^2+1) + r(x)$ where $g(x), r(x) \in F[x]$ with $r(x) = 0$ or $\deg r(x) < 2$. In either case $r(x) = \alpha x + \beta$ with $\alpha, \beta \in F$. Thus $f(x) + \langle x^2+1 \rangle = \alpha x + \beta + \langle x^2+1 \rangle$.

Now α can be chosen in 11 ways and for each choice of α, β can be selected in 11 ways. Hence number of elements of $\frac{F[x]}{\langle x^2+1 \rangle}$ is

121. Similarly it can be proved that $\frac{F[x]}{\langle x^2+x+4 \rangle}$ has 121 elements.

Since each of x^2+1, x^2+x+4 is irreducible, $\langle x^2+1 \rangle$ and $\langle x^2+x+4 \rangle$ are maximal ideals of $F[x]$. Hence $\frac{F[x]}{\langle x^2+1 \rangle}$ and

$\frac{F[x]}{\langle x^2+x+4 \rangle}$ are fields.

Finally define $\theta: \frac{F[x]}{\langle x^2+1 \rangle} \rightarrow \frac{F[x]}{\langle x^2+x+4 \rangle}$ by

$$\theta[\alpha x + \beta + \langle x^2+1 \rangle] = \alpha x + (\beta - 5\alpha) + \langle x^2+x+4 \rangle.$$

It can be checked that θ is a ring homomorphism. Since $\frac{F[x]}{\langle x^2+1 \rangle}$ is a field, θ is 1-1. Also as both fields have same number of elements, θ is onto. Hence $\frac{F[x]}{\langle x^2+1 \rangle} \cong \frac{F[x]}{\langle x^2+x+4 \rangle}$.

Exercise 2. Show that for any integer $n > 1$, the ring $\mathbb{Z}/(n)$ have no non-zero nilpotent elements if and only if n is not divisible by square of any integer.

Solution. Let $\mathbb{Z}/(n)$ has no non-zero nilpotent elements. By Fundamental Theorem of Arithmetic, $n = p_1^{\lambda_1} p_2^{\lambda_2} \dots p_k^{\lambda_k}$ where $\forall i = 1, 2, \dots, k; p_i$'s are distinct primes and $\lambda_i > 0$ are integers. Consider the integer $p_1 p_2 \dots p_k$. Let $\beta = \max(\lambda_1, \lambda_2, \dots, \lambda_k)$. Then $[p_1 p_2 \dots p_k + (n)]^\beta = p_1^\beta p_2^\beta \dots p_k^\beta + (n) = (n)$ as $n \mid p_1^\beta p_2^\beta \dots p_k^\beta$. So we get that $p_1 p_2 \dots p_k + (n)$ is a nilpotent element of $\mathbb{Z}/(n)$. By our assumption this means that $p_1 p_2 \dots p_k + (n) = (n) \Rightarrow n \mid p_1 p_2 \dots p_k$. Hence each $\lambda_i \leq 1$. Since $\lambda_i \geq 1$ for

all i , we have $n = p_1 p_2 \dots p_k$. Consequently n is not divisible by square of any integer.

Conversely let $n = p_1 p_2 \dots p_t$ for distinct primes p_j 's, $j = 1, 2, \dots, t$. Let $m + (n)$ be any nilpotent element of $\mathbb{Z}/(n)$. $[m + (n)]^\alpha = (n)$ for some integer $\alpha \geq 1$.

This implies that $m^\alpha \in (n)$ or $n \mid m^\alpha \Rightarrow p_j \mid m^\alpha \forall j$.

$$\Rightarrow p_j \mid m \forall j \Rightarrow p_1 p_2 \dots p_t \mid m \Rightarrow n \mid m \Rightarrow m + (n) = (n).$$

Hence $\mathbb{Z}/(n)$ has non-zero nilpotent elements.

Exercise 3. If R is a commutative ring with unity, show that centre of the matrix ring R_n consists of all scalar matrices of R_n .

Solution. For convenience sake, we introduce the concept of matrix units E_{ij} ; $i, j = 1, 2, \dots, n$. E_{ij} stands for matrix having zero everywhere except the (i, j) th place, and (i, j) th place has the entry 1, the unity of R . Note that $E_{ij} E_{kl} = 0$ whenever $j \neq k$ and $E_{ik} E_{kl} = E_{il}$. Further every matrix in R_n can be written uniquely as

$$\sum_{j=1}^n \sum_{i=1}^n \alpha_{ij} E_{ij}, \alpha_{ij} \in F.$$

Let $\sum_{j=1}^n \sum_{i=1}^n \alpha_{ij} E_{ij} \in \text{centre } R_n$. Consider $E_{lk} \in R_n$

$$\left(\sum_{j=1}^n \sum_{i=1}^n \alpha_{ij} E_{ij} \right) E_{lk} = E_{lk} \left(\sum_{j=1}^n \sum_{i=1}^n \alpha_{ij} E_{ij} \right)$$

$$\Rightarrow \sum_{i=1}^n \alpha_{il} E_{ik} = \sum_{j=1}^n \alpha_{kj} E_{lj}$$

Now for all $j \neq k$, E_{lj} does not occur on left hand side and so $\alpha_{kj} = 0 \forall j \neq k$. Since E_{lk} was chosen arbitrarily, this shows that all non-diagonal entries of $\sum_{j=1}^n \sum_{i=1}^n \alpha_{ij} E_{ij}$ are zero. Thus each element of

centre of R_n is of the type $\sum_{i=1}^n \alpha_{ii} E_{ii}$.

Again for the matrix $\sum_{k=1}^n \sum_{l=1}^n E_{lk}$

$$\left(\sum_{i=1}^n \alpha_{ii} E_{ii} \right) \left(\sum_{k=1}^n \sum_{l=1}^n E_{lk} \right) = \left(\sum_{k=1}^n \sum_{l=1}^n E_{lk} \right) \left(\sum_{i=1}^n \alpha_{ii} E_{ii} \right)$$

$$\Rightarrow \sum_{k=1}^n \sum_{i=1}^n \alpha_{ii} E_{ik} = \sum_{k=1}^n \sum_{l=1}^n \alpha_{kk} E_{lk}$$

For any fixed k , coefficient of E_{lk} on right hand side is α_{kk} while coefficient of E_{lk} on left hand side is α_{ll} .

$$\text{So } \alpha_{ll} = \alpha_{kk}$$

Since l and k were chosen arbitrarily, we get

$$\alpha_{11} = \alpha_{22} = \dots \alpha_{nn} = \lambda, \text{ say; hence}$$

$$\sum_{i=1}^n \alpha_{ii} E_{ii} = \lambda \sum_{i=1}^n E_{ii}$$

This shows that every element of centre of R_n is a scalar matrix.

Finally let $\lambda \sum_{i=1}^n E_{ii}$ be any scalar matrix in R , then

$$\begin{aligned} \left(\lambda \sum_{i=1}^n E_{ii} \right) \left(\sum_{l=1}^n \sum_{k=1}^n \alpha_{lk} E_{lk} \right) &= \lambda \sum_{k=1}^n \sum_{i=1}^n \alpha_{ik} E_{ik} \\ &= \sum_{k=1}^n \sum_{i=1}^n \lambda \alpha_{ik} E_{ik} = \sum_{k=1}^n \sum_{l=1}^n (\lambda \alpha_{lk}) E_{lk} \\ &\quad \text{(on changing dummy suffix } i \text{ to } l) \end{aligned}$$

$$\text{and } \left(\sum_{l=1}^n \sum_{k=1}^n \alpha_{lk} E_{lk} \right) \left(\lambda \sum_{i=1}^n E_{ii} \right) = \sum_{i=1}^n \sum_{k=1}^n (\alpha_{ik} \lambda) E_{ik}$$

Since R is commutative, $\alpha_{ik} \lambda = \lambda \alpha_{ik}$, hence $\lambda \sum_{i=1}^n E_{ii} \in \text{centre } R_n$.

Thus centre of R_n consists of all scalar matrices in R_n .

Exercise 4. A right ideal A of a ring R is said to be a minimal right ideal of R if (i) $A \neq (0)$ and (ii) for any right ideal B of R , $(0) \subseteq B \subseteq A \Rightarrow$ either $B = (0)$ or $B = A$. If R possesses a minimal right ideal A and for any fixed r if $rA \neq (0)$, show that rA is also a minimal right ideal of R .

Solution. Recall that $rA = \{ra \mid a \in A\}$. So $0 = r0 \in rA$. rA is non-void. For $ra, ra' \in rA$ with $a, a' \in A$, $ra - ra' = r(a - a') \in rA$ as $a - a' \in A$.

For any $x \in R$, $(ra)x = r(ax) \in rA$ as $ax \in A$ (A is right ideal of R).

Thus rA is a right ideal of R . Now by hypothesis $rA \neq (0)$.

Let B be any right ideal of R such that $(0) \subseteq B \subseteq rA$. We shall prove that if $B \neq (0)$ then $B = rA$.

Since $B \neq (0)$, there exists $b (\neq 0) \in B$. As $B \subseteq rA$, $b = ra$ for some $a \in A$. Let $X = \{s \in A \mid rs \in B\}$.

Clearly $0 \in X$, X is non-empty. Further $a_1, a_2 \in X \Rightarrow ra_1, ra_2 \in B \Rightarrow ra_1 - ra \in B \Rightarrow r(a_1 - a_2) \in B \Rightarrow a_1 - a_2 \in X$. Again for any $x \in R$, $ra_1 \in B \Rightarrow (ra_1)x \in B$ as B is a right ideal of R . This implies that $r(a_1 x) \in B \Rightarrow a_1 x \in X$.

Hence X is a right ideal of R , contained in A ; further since $a \in X$, $X \neq (0)$. By minimality of A , this forces X to be equal to A . In other words we get that $rA \subseteq B$. Hence $B = rA$.

Consequently rA is a minimal right ideal of R .

Exercise 5. Let M be a proper ideal in a Boolean ring R (i.e. $x^2=x$ for all $x \in R$) with unity. Prove that (i) R/M is a Boolean ring and (ii) $R/M \cong \mathbb{Z}/(2)$ if and only if M is a maximal ideal in R .

Solution. (i) For any $x+M \in R/M$, $(x+M)^2 = x^2+M = x+M$ as $x = x$ in R . Hence $\frac{R}{M}$ is a Boolean ring.

(ii) If $\frac{R}{M} \cong \frac{\mathbb{Z}}{(2)}$, then since $\frac{\mathbb{Z}}{(2)}$ is a field, M is a maximal ideal of R .

Conversely let M be a maximal ideal of R , then R/M is a field and since $(x+M)^2 = x+M$ for all $x \in R$, we get

$(x+M)[(x+M)-(1+M)] = M \Rightarrow x+M = M$ or $x+M = 1+M$ as R/M is a field. Thus R/M has only two elements namely M and $1+M$.

The mapping $f: \frac{R}{M} \rightarrow \frac{\mathbb{Z}}{(2)}$ given by $f(M) = (2)$ and $f(1+M) = 1+(2)$ is clearly an isomorphism.

Hence $R/M \cong \mathbb{Z}/(2)$.

MISCELLANEOUS PROBLEMS

1. Let R be a ring with unity. Prove that the sets $A = \{x \in R \mid x \text{ is not invertible}\}$, $B = \{x \in R \mid x \text{ is not left invertible}\}$, $C = \{x \in R \mid x \text{ is not right invertible}\}$ are all equal if any one of them is closed under addition.

2. Prove that a ring R is Boolean if and only if it has no non zero nilpotent element and $(a+b)ab=0$ for all $a, b \in R$.

3. If m and n are relatively prime integers, show that the ring $\mathbb{Z}/(mn)$ contains at least two idempotent elements other than the zero and unity.

4. Prove that the ring $\mathbb{Z}/(n)$ has no idempotent other than zero and unity if and only if $n = p^k$ for some prime number p and integer $k \geq 1$.

5. Show that a UFD in which every proper prime ideal is maximal, is a PID.

6. A commutative ring R is called a *valuation ring* if for every pair of non-zero elements x, y of R , either $x \mid y$ or $y \mid x$. Show that every such ring has unity.

7. An integral domain with unity is said to be a *factorization domain* if every non-zero non-unit element of R can be written as $p_1 p_2 \dots p_k$ where p_i 's are irreducible elements of R and k is a natural

number. Prove (i) A factorization domain R is a UFD if and only if every irreducible element of R is prime. (ii) A factorization domain R is a UFD if and only if every pair of non-zero elements has an HCF.

8. If F is a field and $p(x)$, a non-zero non-unit element of $F[x]$, prove that the ring $R = \frac{F[x]}{\langle p[x] \rangle}$ has no non-zero nilpotent elements if and only if $p(x)$ is not divisible by the square of any polynomial.

[Hint. Imitate the solution of Exercise 2 of miscellaneous worked-out exercises.]

9. A ring R is said to be *Regular* (in the sense of Von-Neumann) if for each $x \in R$, there exists y in R such that $xyx = x$. Prove (i) If I is an ideal of a regular ring R , then every ideal of I is an ideal of R . (ii) If R is a commutative regular ring with unity 1 and 1 is the only unity of R , then R is a Boolean ring. (iii) If R is a Boolean ring with unity 1 then R is a regular ring and 1 is the only unit in R .

10. Show that in a Boolean ring R with unity, every prime ideal different from R is a maximal ideal

[Hint. Let P be a prime ideal of R such that $P \neq R$. Let $x \in R - P$, then $x' = x \Rightarrow x(1-x) = 0 \in P \Rightarrow 1-x \in P \Rightarrow 1 \in P + xR$.]

11. An ideal A of a ring R is called *Modular* if the ring R/A has unity. Prove that

(i) Every ideal of a ring with unity is modular.

(ii) Every ideal containing a modular ideal is modular.

(iii) The intersection of two modular ideals is modular.

[Hint. (iii) If $\bar{e}$ and $\bar{f}$ are unities of R/A and R/B respectively, then $e+f-ef$ is unity of $R/A \cap B$.]

12. Prove that an ideal A of a ring R is modular if and only if there exists $e \in R$ such that $ex = x$ and $x - ex \in A$ for all $x \in A$.

13. If f is a homomorphism from R onto a field F , prove that $\text{Ker } f$ is a maximal ideal of R .

14. (Kaplansky). Let a be nilpotent element of a ring R . Let b be a left zero-divisor of R . Suppose a and b commute. Prove that there exists a $c \neq 0$ in R such that $ac = bc = 0$.

15. Let R be a ring with unity 1. Show that if $1+1$ is a unit, then any idempotent is the sum of two units.

16. Let R be a ring having at least two elements. Show that R is a field if and only if for all a, b, c in R with $a \neq b$, the equation $ax - xb = c$ has a unique solution in R .

17. Show that if R is a Von-Neumann regular ring with unity, then following statements are equivalent:

(i) Every principal left ideal is generated by an idempotent.

(ii) For every principal left ideal Ra of R , there exists $b \in R$ such that $R = Ra \oplus Rb$.

[Hint. A ring R is said to be *direct sum* of subrings X and Y if $R = X + Y$ and $X \cap Y = (0)$. X and Y are called *direct summand* of R , we write $R = X \oplus Y$.]

(iii) Every principal left ideal of R is a direct summand of R .

(iv) Every finitely generated left ideal is a direct summand of R .

18. Prove that in a Von-Neumann regular ring R , every finitely generated left ideal is principal and intersection of any two principal left ideals of R is a principal left ideal of R .

19. If R is a Von-Neumann regular ring with unity and if it has no non-zero nilpotent elements; show that every one-sided ideal of R is a two-sided ideal of R .

20. If P be the set of non-zero positive real numbers, define $*$ on P by $x * y = x^{\log_2 y} \forall x, y \in P$. Prove that $\langle P, *, \cdot \rangle$ is a field isomorphic to $\mathbf{R}$.

11

VECTOR SPACES

In Analytical Geometry and Mechanics, one comes across the concept of a vector as a directed magnitude, we do not need recall that definition here. From algebraic point of view the vectors so defined have the following salient features.

1. In three-dimensional Euclidean space, a vector v is determined uniquely by its three components (ξ, η, λ) (relative to a definite co-ordinate system) and conversely given any ordered triple (ξ, η, λ) of real numbers there exists a vector v having ξ, η, λ as its co-ordinates. It is customary to write $v = (\xi, \eta, \lambda)$, to emphasize that v is a vector whose components are ξ, η and λ . Similarly in the two-dimensional Euclidean space, a vector has two co-ordinates and is determined uniquely by its co-ordinates. Further the real numbers are called *scalars*.

2. There are three fundamental operations on vectors, namely addition of vectors, the multiplication of a vector by a scalar and the scalar product of vectors. Thus if $v = (\xi, \eta, \lambda)$ and $v' = (\xi', \eta', \lambda')$ are two vectors in $\mathbf{R}^3$, the three-dimensional Euclidean space, α is any scalar, then we define $v + v' = (\xi + \xi', \eta + \eta', \lambda + \lambda')$, $\alpha v = (\alpha\xi, \alpha\eta, \alpha\lambda)$ and scalar product $(v, v') = \xi\xi' + \eta\eta' + \lambda\lambda'$. First two operations on vectors play very significant part in Geometry and Mechanics. In this chapter we endeavour to give a treatment of vectors in a more general setting.

The study of vectors relative to addition and multiplication by real numbers can be generalized in two directions. Firstly it is not necessary for us to restrict our study to ordered pairs or ordered triples, instead one may take ordered n -tuples. Secondly we do not need confine ourselves to co-ordinates being real numbers. The co-ordinates may be taken from any field or more generally from any division ring. We now give an abstract definition of a vector space and then see that how does it generalize the above mentioned concept of vectors in Analytical Geometry and Mechanics.

1. Definition and Elementary Properties

Definition 11.1. A system $(V, D, \oplus, \cdot)$ where V is a non-void set, D a division ring, $\oplus$ a binary composition on V and $\cdot$ is a mapping of $D \times V$ into V which for each element $v \in V$, $\alpha \in D$ determines an element $\alpha \cdot v \in V$, is called a *left vector space* over D if it satisfies the following conditions.

$V-1.$ $\langle V, \oplus \rangle$ is an Abelian group.

$V-2.$ For all $x, y \in V$, $\alpha, \beta \in D$ we have

$$(i) \quad \alpha \cdot (x \oplus y) = \alpha \cdot x \oplus \alpha \cdot y$$

$$(ii) \quad (\alpha + \beta) \cdot x = \alpha \cdot x \oplus \beta \cdot x$$

$$(iii) \quad (\alpha\beta) \cdot x = \alpha \cdot (\beta \cdot x)$$

$$(iv) \quad 1 \cdot x = x.$$

Every element of V is called a *vector*, every member of D is called a *scalar*.

Similarly if in the above definition we replace $\alpha \cdot v$ by $v \cdot \alpha$, then we get a *right vector space* over D .

REMARK. In analogy with our convention in groups and rings, when there is no likelihood of any confusion we shall say that V is a vector space over a division ring D , whenever the system $(V, D, \oplus, \cdot)$ is a vector space over D . In this book we are interested only in studying vector spaces over fields. However we emphasize that most of the results proved here for vector spaces over fields can be suitably generalized to vector spaces over division rings.

Note. For convenience sake, in future, same symbol '+' will stand for $\oplus$ on V and addition in D .

EXAMPLE 1. For any field F , let $V = \{(\alpha_1, \alpha_2) \mid \alpha_i \in F\}$. Then V is a vector space over F under vector addition and multiplication of a vector by scalar given by

$$(\alpha_1, \alpha_2) + (\beta_1, \beta_2) = (\alpha_1 + \beta_1, \alpha_2 + \beta_2)$$

$$\alpha(\alpha_1, \alpha_2) = (\alpha\alpha_1, \alpha\alpha_2).$$

EXAMPLE 2. Let V be the additive group of polynomials in one variable over a field F . For any $\alpha \in F$ and $f(x) = \alpha_0 + \alpha_1 x + \dots + \alpha_n x^n \in F[x]$, define $\alpha f(x) = \alpha\alpha_0 + \alpha\alpha_1 x + \dots + \alpha\alpha_n x^n$. Then V becomes a vector space over F .

EXAMPLE 3. Let V be the set of all continuous real-valued functions defined on the closed interval $[0, 1]$. For any $f, g \in V$ and $\alpha \in \mathbb{R}$, Define $f+g$ and αf by

$$(f+g)(x) = f(x) + g(x)$$

$$(\alpha f)(x) = \alpha f(x) \quad \forall x \text{ in } [0, 1].$$

Under these operations, V becomes a vector space over $\mathbb{R}$ (since the sum of continuous functions and scalar multiple of any continuous function are continuous).

EXAMPLE 4. For any field F and any positive integer n , the set of all n -tuples $(\alpha_1, \alpha_2, \dots, \alpha_n)$ of elements of F , is a vector space over F under vector addition and multiplication of a vector by a scalar given by

$$(\alpha_1, \alpha_2, \dots, \alpha_n) + (\beta_1, \beta_2, \dots, \beta_n) = (\alpha_1 + \beta_1, \alpha_2 + \beta_2, \dots, \alpha_n + \beta_n)$$

$$\alpha(\alpha_1, \alpha_2, \dots, \alpha_n) = (\alpha\alpha_1, \alpha\alpha_2, \dots, \alpha\alpha_n).$$

This vector space is usually denoted by $F^{(n)}$ or by F^n .

As remarked earlier, we are only interested here in studying vector spaces over fields. Following lemma shows that there is no distinction between right and left vector spaces over a field.

Lemma 11.2. *If V is a left vector space over a field F then V is also a right vector space over F and conversely.*

Proof: Since V is a left vector space over F , for each $v \in V$ and $\alpha \in F$, $\alpha.v$ is a uniquely defined element of V such that

$$\alpha.(u+v) = \alpha.u + \alpha.v$$

$$(\alpha + \beta).u = \alpha.u + \beta.u$$

$$(\alpha\beta).u = \alpha.(\beta.u)$$

and

$$1.u = u \quad \forall \alpha, \beta \in F, u, v \in V.$$

Define $v.\alpha$ as $\alpha.v \quad \forall \alpha \in F, v \in V$.

Then for any $\alpha, \beta \in F, u, v \in V$, we have

$$(u+v).\alpha = \alpha.(u+v) = \alpha.u + \alpha.v = u.\alpha + v.\alpha.$$

$$u.(\alpha + \beta) = (\alpha + \beta).u = \alpha.u + \beta.u = u.\alpha + u.\beta$$

$$u.(\alpha\beta) = (\alpha\beta).u = (\beta\alpha).u \quad \text{as } \alpha\beta = \beta\alpha$$

$$= \beta.(\alpha.u) = \beta.(u.\alpha) = (u.\alpha).\beta$$

and

$$u.1 = 1.u = u.$$

Thus V is an Abelian group under addition such that for every $v \in V, \alpha \in F, v.\alpha$ is a uniquely defined element of V satisfying the following conditions:

For any $u, v \in V, \alpha, \beta \in F$

$$(u+v).\alpha = u.\alpha + v.\alpha$$

$$u.(\alpha + \beta) = u.\alpha + u.\beta$$

$$u.(\alpha\beta) = (u.\alpha).\beta$$

$$u.1 = u.$$

Hence V is a right vector space over F .

Similarly if V is a right vector space over F then V is also a left vector space over F . This proves the lemma. ■

Henceforth we shall confine ourselves to vector spaces over fields; therefore we shall not make any distinction between left and right vector spaces. However as far as possible we shall continue to write the scalars on the left in the multiplication of a scalar and a vector. that V is a vector space over a field F .

NOTATION. V_F will denote that V is a vector space over a field F .

REMARK 1. For the sake of convenience, in future, we shall write

αv in place of α, v .

REMARK 2. We give below an example to show that the condition $V-2$ (iv) viz. $1.x = x \forall x \in V$ is independent of the remaining conditions $V-1$ and $V-2$ (i), (ii), (iii).

EXAMPLE 5. Consider $V = \{(\alpha, \beta, \gamma) \mid \alpha, \beta, \gamma \in \mathbf{R}\}$.

Define $(\alpha, \beta, \gamma) + (\alpha', \beta', \gamma') = (\alpha + \alpha', \beta + \beta', \gamma + \gamma')$

$$\lambda(\alpha, \beta, \gamma) = (\lambda\alpha, \lambda\beta, 0) \forall \alpha, \beta, \gamma, \alpha', \beta', \gamma', \lambda \in \mathbf{R}.$$

Now it can be checked that all conditions $V-1, V-2$ (i), (ii) and (iii) are satisfied.

Now $(3, 2, 6) \in V$, but $1(3, 2, 6) = (3, 2, 0) \neq (3, 2, 6)$.

Hence V is not a vector space over $\mathbf{R}$.

Lemma 11.3. Let V be a vector space over a field F and let O_V and O_F be the additive identities of V and F respectively, then the following hold.

- (1) $\alpha O_V = O_V \forall \alpha \in F$.
- (2) $O_F v = O_V \forall v \in V$.
- (3) $-v = (-1)v \forall v \in V$.
- (4) $(-\alpha)v = \alpha(-v) = -(\alpha v) \forall \alpha \in F, v \in V$.
- (5) If $\alpha v = O_V$ then either $\alpha = O_F$ or $v = O_V$.

Proof: (1) Since $O_V = O_V + O_V$ we have

$$\alpha O_V = \alpha(O_V + O_V) = \alpha O_V + \alpha O_V \quad [V-2]$$

$\Rightarrow \alpha O_V = O_V$ by cancellation in the group $\langle V, + \rangle$.

- (2) $O_F = O_F + O_F$ gives $O_F v = O_F v + O_F v$

$\Rightarrow O_F v = O_V$ by cancellation in the group $\langle V, + \rangle$.

- (3) Now $[1 + (-1)]v = 1.v + (-1)v$.

Also $O_V = O_F v = [1 + (-1)]v = 1v + (-1)v = v + (-1)v$.

Hence $(-1)v = -v$.

- (4) Since $O_V = O_F v = [\alpha + (-\alpha)]v = \alpha v + (-\alpha)v$ we get $(-\alpha)v = -(\alpha v)$.

Since $O_V = \alpha O_V = \alpha[v + (-v)] = \alpha v + \alpha(-v)$

we also get $\alpha(-v) = -(\alpha v)$.

Hence $(-\alpha)v = \alpha(-v) = -(\alpha v)$.

- (5) Let $\alpha v = O_V$. Suppose $\alpha \neq O_F$, then α^{-1} exists in F and $v = 1v = (\alpha^{-1}\alpha)v = \alpha^{-1}O_V = O_V$.

Hence either $\alpha = O_F$ or $v = O_V$.

Note. In future we shall use the same symbol O or O for the zero vector O_V and for the zero scalar O_F , for any vector space V_F . It should be clear to the readers from the context for which thing O or O is being used at a particular occasion.

2. Subspaces

Definition 11.4. A non-void subset W of a vector space V_F is said to be a subspace of V if

S-1. For any $a, b \in W$, $a+b \in W$

S-2. For any $a \in W$ and $\alpha \in F$, $\alpha a \in W$.

It is immediate from the definition that (0) and V are both subspaces of V . These are trivial subspaces. Any subspace W of V_F which is different from (0) and V is called a *proper subspace* of V .

Lemma 11.5. *If W is a subspace of a vector space V_F , then W is a subgroup of $\langle V, + \rangle$ and W is a vector space over F under induced addition and multiplication of vectors by scalars.*

Proof: For any $x, y \in W$, we have $(-1)y \in W$ by S-2; so by S-1, $x + (-1)y \in W \Rightarrow x - y \in W$.

This proves that W is a subgroup of $\langle V, + \rangle$.

The property V-2 in definition 11.1 is satisfied by elements of W , since $W \subseteq V$ and the same property holds for V and also the compositions in W are all induced by those in V . Hence W is a vector space over F under induced addition and the multiplication of vectors by scalars. ■

EXAMPLE 6. Consider $\mathbf{R}^3 = \{(a_1, a_2, a_3) \mid a_i \in \mathbf{R}\}$.

Let $W_1 = \{(0, a_2, a_3) \mid a_2, a_3 \in \mathbf{R}\}$

$W_2 = \{(a_1, a_2, a_3) \mid a_1 + a_2 + a_3 = 0\}$.

It can be easily seen that W_1 and W_2 both satisfy the conditions which define a subspace. Consequently W_1, W_2 are both subspaces of $\mathbf{R}^3$.

EXAMPLE 7. Let $\mathbf{C}$ be the set of all complex numbers, $\mathbf{C}$ is a vector space over $\mathbf{R}$, the field of real numbers. $W = \{ib \mid b \in \mathbf{R}, i = \sqrt{-1}\}$ is a subspace of $\mathbf{C}$.

Theorem 11.6. *A non-void subset W of a vector space V_F is a subspace of V if and only if it satisfies the following condition:*

$\alpha, \beta \in F; a, b \in W$ implies $\alpha a + \beta b \in W$.

Proof: Let W be a subspace of V and let $a, b \in W, \alpha, \beta \in F$. Then $\alpha a \in W, \beta b \in W$ (Definition 11.4, S-2). This further gives $\alpha a + \beta b \in W$ (Definition 11.4, S-1).

Conversely let W satisfy the given condition. Let $a, b \in W$. Taking $\alpha = 1, \beta = 1$, we get $a + b = 1a + 1b \in W$. Again taking $\beta = 0$ and α to be arbitrary we get $\alpha a = \alpha a + 0b \in W$. Hence W is a subspace of V . ■

Theorem 11.7. *Intersection of any family of subspaces of a vector space is a subspace.*

Proof: Let $\{W_\lambda \mid \lambda \in \Lambda\}$ be any family of subspaces of a vector space V_F . Let $W = \bigcap_{\lambda \in \Lambda} W_\lambda$.

Since $0 \in W_\lambda \forall \lambda \in \Lambda$, we have $0 \in W$. So W is a non-void subset of V .

Now $a, b \in W, \alpha, \beta \in F$

$$\Rightarrow a \in W_\lambda, b \in W_\lambda \quad \forall \lambda \in \Lambda, \text{ as } W = \bigcap_{\lambda \in \Lambda} W_\lambda.$$

$\Rightarrow \alpha a + \beta b \in W_\lambda \quad \forall \lambda \in \Lambda$, since each W_λ is a subspace of V (Theorem 11.6).

$$\Rightarrow \alpha a + \beta b \in W.$$

Hence W is a subspace of V . ■

In particular above theorem shows that intersection of any two subspaces of a vector space V_F is a subspace of V_F . We proved before that the union of two subgroups need not be a subgroup, the analogous result holds for subspaces.

EXAMPLE 8. Consider $\mathbf{R}^2$. Then

$$W_1 = \{(a, 0) \mid a \in \mathbf{R}\}$$

$$W_2 = \{(0, b) \mid b \in \mathbf{R}\}$$

are subspaces of $\mathbf{R}^2$. Let $W = W_1 \cup W_2$. Now $(1, 0) \in W_1$, $(0, 1) \in W_2$ but $(1, 1)$ is neither in W_1 nor in W_2 . So that $(1, 1) = (1, 0) + (0, 1) \notin W$. Hence W is not closed under vector addition. Consequently $W_1 \cup W_2$ is not a subspace of $\mathbf{R}^2$.

Definition 11.8. Let X be a subset of a vector space V_F then a subspace W of V is said to be *spanned by* or *generated by* X if

$$(i) \quad X \subseteq W$$

$$(ii) \quad \text{For any subspace } W' \text{ of } V, X \subseteq W' \Rightarrow W \subseteq W'.$$

Notation. $\langle X \rangle$ will denote the subspace spanned by X . Further if X consists of elements $x_1, x_2, \dots, x_n$ then we simply write $\langle X \rangle = \langle x_1, x_2, \dots, x_n \rangle$.

Theorem 11.9. For any finite non-void subset $X = \{x_1, x_2, \dots, x_n\}$, the subspace spanned by X is the set of all vectors of the form $\alpha_1 x_1 + \alpha_2 x_2 + \dots + \alpha_n x_n$ where $\alpha_i \in F$ for $i = 1, 2, \dots, n$.

Proof: Let W be the set of all vectors in V of the form $\alpha_1 x_1 + \alpha_2 x_2 + \dots + \alpha_n x_n$, $\alpha_i \in F, i = 1, 2, \dots, n$. Taking $\alpha_1 = 1, \alpha_i = 0 \forall i \geq 2$, we see that $x_1 \in W$. Similarly $x_2, x_3, \dots, x_n$ are all in W . So that $X \subseteq W$. Suppose W' is a subspace of V such that $X \subseteq W'$. Then for each i , $\alpha_i \in F, x_i \in X \subseteq W' \Rightarrow \alpha_i x_i \in W'$. This gives $\alpha_1 x_1 + \alpha_2 x_2 + \dots + \alpha_n x_n \in W'$. Thus $W \subseteq W'$. So we see that W is a subset of V satisfying conditions (i) and (ii) in definition 11.8. So the result will follow if we show that W is a subspace of V . To see this let

$$a = \alpha_1 x_1 + \alpha_2 x_2 + \dots + \alpha_n x_n, b = \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n$$

be any two elements of W , further let $\alpha, \beta \in F$ then $\alpha a + \beta b = (\alpha \alpha_1 + \beta \beta_1) x_1 + (\alpha \alpha_2 + \beta \beta_2) x_2 + \dots + (\alpha \alpha_n + \beta \beta_n) x_n \in W$. Hence W is a subspace of V . This proves the theorem. ■

Definition 11.10. For any finite number of vectors $x_1, x_2, \dots, x_n$ in a vector space V_F and scalars $\alpha_1, \alpha_2, \dots, \alpha_n$ in F , the vector $\alpha_1 x_1 + \alpha_2 x_2 + \dots + \alpha_n x_n$ is called a *linear combination* of $x_1, x_2, \dots, x_n$.

Thus, Theorem 11.9 gives that if X is a finite non-void subset of V_F , the subspace $\langle X \rangle$ is the set of all linear combinations of members of X .

Definition 11.11. For any two subspaces W_1 and W_2 of a vector space V_F , their sum $W_1 + W_2$ is the set given by $\{w_1 + w_2 \mid w_1 \in W_1, w_2 \in W_2\}$.

Theorem 11.12. For any two subspaces W_1 and W_2 of a vector space V_F , $W_1 + W_2$ is the subspace of V spanned by $W_1 \cup W_2$.

Proof: Firstly we show that $W_1 + W_2$ is a subspace of V . Clearly $0 = 0 + 0 \in W_1 + W_2$ so $W_1 + W_2 \neq \emptyset$. Let $x, y \in W_1 + W_2$, $\alpha, \beta \in F$. Then $x = w_1 + w_2$, $y = w_1' + w_2'$, $w_1, w_1' \in W_1$; $w_2, w_2' \in W_2$.

$$\begin{aligned} \text{Thus } \alpha x + \beta y &= (\alpha w_1 + \alpha w_2) + (\beta w_1' + \beta w_2') \\ &= (\alpha w_1 + \beta w_1') + (\alpha w_2 + \beta w_2') \end{aligned} \quad \dots (1)$$

Now $\alpha w_1 + \beta w_1' \in W_1$, $\alpha w_2 + \beta w_2' \in W_2$.

Consequently (i) gives that $\alpha x + \beta y \in W_1 + W_2$.

Hence $W_1 + W_2$ is a subspace of V .

Now we show that $W_1 + W_2$ is spanned by $W_1 \cup W_2$. For any $w_1 \in W_1$ as $0 \in W_2$ we have $w_1 = w_1 + 0 \in W_1 + W_2$.

So that $W_1 \subseteq W_1 + W_2$. Similarly $W_2 \subseteq W_1 + W_2$ (2)

Consequently $W_1 \cup W_2 \subseteq W_1 + W_2$.

Let W be any subspace of V containing $W_1 \cup W_2$. Then given $w_1 \in W_1$, $w_2 \in W_2$, we have $w_1, w_2 \in W_1 \cup W_2 \Rightarrow w_1, w_2 \in W \Rightarrow w_1 + w_2 \in W$ (definition since W is closed under vector addition $\Rightarrow W_1 + W_2 \subseteq W$ (definition) ... (3)

11.11).

From (2) and (3) we get that $W_1 \cup W_2$ spans $W_1 + W_2$. ■

EXAMPLE 9. For any field F , consider the vector space $F^{(3)}$. $e_1 = (1, 0, 0)$, $e_2 = (0, 1, 0)$ and $e_3 = (0, 0, 1)$ are in $F^{(3)}$. Let $x = (\alpha_1, \alpha_2, \alpha_3) \in F^{(3)}$. Then $x = \alpha_1(1, 0, 0) + \alpha_2(0, 1, 0) + \alpha_3(0, 0, 1) = \alpha_1 e_1 + \alpha_2 e_2 + \alpha_3 e_3$. Hence every member of $F^{(3)}$ is a linear combination of e_1, e_2, e_3 . Consequently $\{e_1, e_2, e_3\}$ spans $F^{(3)}$. In general if we consider $F^{(n)}$, and define $e_1 = (1, 0, 0, \dots, 0)$, $e_2 = (0, 1, 0, \dots, 0)$, ..., $e_n = (0, 0, 0, \dots, 1)$, then F^n is spanned by the n vectors $e_1, e_2, \dots, e_n$.

EXAMPLE 10. Consider $\mathbb{R}^3$. $x_1 = (1, 1, 0)$, $x_2 = (1, 2, 0) \in \mathbb{R}^3$. Let W be the subspace spanned by $\{x_1, x_2\}$. Then $x_1, x_2 \in W \Rightarrow x_2 - x_1 \in W \Rightarrow (0, 1, 0) \in W \Rightarrow (1, 0, 0) = x_1 - (0, 1, 0) \in W$. Thus $e_1 = (1, 0, 0)$, $e_2 = (0, 1, 0) \in W \Rightarrow$ for every $\alpha_1, \alpha_2 \in \mathbb{R}$, $\alpha_1 e_1 + \alpha_2 e_2 = (\alpha_1, \alpha_2, 0) \in W$.

Let $W' = \{(\alpha_1, \alpha_2, 0) \mid \alpha_1, \alpha_2 \in \mathbb{R}\}$. W' is a subspace of $\mathbb{R}^3$ and as seen above $W' \subseteq W$. Now $x_1 = e_1 + e_2 \in W'$, $x_2 = e_1 + 2e_2 \in W'$ since $e_1, e_2 \in W'$. Thus the subspace $W = \langle x_1, x_2 \rangle \subseteq W'$. Hence $W = W'$.

WORKED-OUT EXERCISES

Exercise 1. Let F be a field and V , a non-empty set with a binary composition $\oplus$. If for each $\alpha \in F$, $v \in V$ a unique element $\alpha \times v$ of V is defined satisfying the following axioms. Then show that $\langle V, \oplus, \times \rangle$ is a vector space over F .

- (1) $(a \oplus b) \oplus c = a \oplus (b \oplus c) \quad \forall a, b, c \in V$.
- (2) $\lambda \times (a \oplus b) = \lambda \times a \oplus \lambda \times b \quad \forall \lambda \in F, a, b \in V$.
- (3) $(\lambda + \mu) \times a = \lambda \times a \oplus \mu \times a \quad \forall \lambda, \mu \in F, a \in V$.
- (4) $(\lambda \mu) \times a = \lambda \times (\mu \times a) \quad \forall \lambda, \mu \in F, a \in V$.
- (5) $0 \times a = 0 \times b \quad \forall a, b \in V$.
- (6) $1 \times a = a \quad \forall a \in V$.

Solution. Put $0 \times a = z$. Then because of (5), z is independent of a . For any $v \in V$, $v = 1 \times v$ by (6)

$$\begin{aligned} &= (1 + 0) \times v \\ &= 1 \times v \oplus 0 \times v \quad \text{by (3)} \\ &= v \oplus z. \end{aligned}$$

Thus z is right identity of V for the composition $\oplus$. Further for every $v \in V$,

$$\begin{aligned} z &= 0 \times v = [1 + (-1)] \times v = 1 \times v \oplus (-1) \times v \\ &= v \oplus (-1) \times v \end{aligned}$$

This shows that $(-1) \times v$ is right inverse of v . This in turn implies that $\langle V, \oplus \rangle$ is group under $\oplus$.

Now $(1+1) \times (a \oplus b) = 1 \times (a \oplus b) \oplus 1 \times (a \oplus b) \quad \text{by (3)}$

$$\begin{aligned} &= [1 \times a \oplus 1 \times b] \oplus [1 \times a \oplus 1 \times b] \quad \text{by (2)} \\ &= a \oplus b \oplus a \oplus b. \quad \text{by (6)} \end{aligned}$$

Again $(1+1) \times (a \oplus b) = (1+1) \times a \oplus (1+1) \times b \quad \text{by (2)}$

$$\begin{aligned} &= 1 \times a \oplus 1 \times a \oplus 1 \times b \oplus 1 \times b \quad \text{by (3)} \\ &= a \oplus a \oplus b \oplus b \quad \text{by (6)} \end{aligned}$$

Thus $a \oplus b \oplus a \oplus b = a \oplus a \oplus b \oplus b$.

By right and left cancellation laws in group $\langle V, \oplus \rangle$, we get $b \oplus a = a \oplus b$.

Hence $\langle V, \oplus \rangle$ is an Abelian group.

Consequently $\langle V, \oplus, \times \rangle$ is a vector space over F .

Exercise 2. Let $\mathbf{R}$ be the field of all real numbers and let V be the set of all infinite sequences $\langle a_n \rangle$, $a_n \in \mathbf{R}$, where equality, addition and scalar multiplication are defined term wise. Prove that V is a vector space over $\mathbf{R}$ and $U = \{ \langle x_n \rangle \mid \sum x_n^2 \text{ is convergent} \}$ is a subspace of V .

Solution. It can be easily verified that V is a vector space over $\mathbf{R}$. We show that U is a subspace of V .

Clearly the sequence $\langle x_n \rangle$ with $x_n = 0 \quad \forall n \in \mathbf{N}$, is in U , so U is non-empty.

Let $\langle x_n \rangle, \langle y_n \rangle \in U$, then $\sum x_n^2$ and $\sum y_n^2$ are convergent series.

Thus $\lim_{n \rightarrow \infty} \sum_{i=1}^n x_i^2 = L$ and $\lim_{n \rightarrow \infty} \sum_{i=1}^n y_i^2 = M$ for some real numbers

L and M .

$$\begin{aligned} \text{Now } (x_i - y_i)^2 &= x_i^2 + y_i^2 - 2x_i y_i = 2x_i^2 + 2y_i^2 - (x_i + y_i)^2 \\ &\leq 2x_i^2 + 2y_i^2 \text{ as } (x_i + y_i)^2 \geq 0. \end{aligned}$$

$$\text{This gives that } \sum_{i=1}^n (x_i - y_i)^2 \leq 2 \sum_{i=1}^n x_i^2 + 2 \sum_{i=1}^n y_i^2$$

$$\begin{aligned} \text{Consequently } \lim_{n \rightarrow \infty} \sum_{i=1}^n (x_i - y_i)^2 &\leq 2 \lim_{n \rightarrow \infty} \sum_{i=1}^n x_i^2 + 2 \lim_{n \rightarrow \infty} \sum_{i=1}^n y_i^2 \\ &= 2L + 2M \end{aligned}$$

$\Rightarrow \sum (x_i - y_i)^2$ is a convergent series.

As a result of this we get that $\langle x_n \rangle - \langle y_n \rangle = \langle x_n - y_n \rangle \in U$.

$$\text{Finally for any } \alpha \in \mathbf{R}, \lim_{n \rightarrow \infty} \sum_{i=1}^n (\alpha x_i)^2 = \alpha^2 L$$

$\Rightarrow \sum (\alpha x_i)^2$ is convergent $\Rightarrow \alpha \langle x_n \rangle = \langle \alpha x_n \rangle \in U$.

Hence U is a subspace of V .

Exercise 3. If V is a vector space over an infinite field prove that V cannot be written as set-theoretic union of a finite number of proper subspaces.

Solution. Let if possible, $V = \bigcup_{i=1}^n V_i$. We can assume without loss of generality that this union is irredundant, i.e.

$$V_i \not\subseteq \bigcup_{\substack{j=1 \\ j \neq i}}^n V_j \text{ for all } i.$$

$$\text{This implies } \bigcup_{i=1}^{n-1} V_i \not\subseteq V_n \text{ and } V_n \not\subseteq \bigcup_{i=1}^{n-1} V_i.$$

$$\text{There exist } v \in V_n, v \notin \bigcup_{i=1}^{n-1} V_i \text{ and } w \in \bigcup_{i=1}^{n-1} V_i, w \notin V_n.$$

Consider the set $X = \{\lambda v + w \mid \lambda \in F\}$. No element of X is in V_n , otherwise $w \in V_n$. But $V = \bigcup_{i=1}^n V_i \Rightarrow$ all elements of X are in $\bigcup_{i=1}^{n-1} V_i$.

Since F is an infinite field, we can find $1 \leq i_0 \leq n-1$ and two distinct elements α and β of F such that $\alpha v + w \in V_{i_0}$, $\beta v + w \in V_{i_0}$ for some $1 \leq i_0 \leq n-1$.

$$\text{This in turn implies that } (\alpha v + w) - (\beta v + w) \in V_{i_0} \subseteq \bigcup_{i=1}^{n-1} V_i$$

$$\begin{aligned} \Rightarrow (\alpha - \beta)v &\in \bigcup_{i=1}^{n-1} V_i \\ \Rightarrow (\alpha - \beta)^{-1}(\alpha - \beta)v &\in \bigcup_{i=1}^{n-1} V_i \quad \text{as } \alpha \neq \beta, \\ \Rightarrow v &\in \bigcup_{i=1}^{n-1} V_i. \end{aligned}$$

This is against the choice of v .

Hence $V = \bigcup_{i=1}^n V_i$ is impossible.

Exercise 4. An Abelian group $\langle V, + \rangle$ is called **divisible** if for any non-zero integer n , $nV = V$, i.e. if for every $u \in V$, and for any non-zero integer n , there exists $v \in V$ such that $u = nv$. Show that $\langle V, + \rangle$ is a vector space over $\mathbf{Q}$, if and only if V is a divisible group, all of whose non-zero elements are of infinite order.

Solution. Let V be a divisible group all of whose non-zero elements are of infinite order.

For any integer $q \neq 0$, and $u \in V$, there exists $v \in V$ such that $u = qv$. Further if $u = qv = qv'$ then $q(v - v') = 0$. But this forces $v = v'$ since V has no non-zero element of finite order. For any $\frac{p}{q} \in \mathbf{Q}$ and $u \in V$, define $\frac{p}{q} u = pv$ where $u = qv$.

This is well-defined, since for each $u \in V$, and $q \neq 0 \in \mathbf{Z}$ there exists unique v such that $u = qv$.

It is routine matter to check that $\left(\frac{p}{q} + \frac{r}{s}\right)u = \frac{pu}{q} + \frac{ru}{s}$, $\frac{p}{q}(u+v) = \frac{pu}{q} + \frac{pv}{q}$ and $1u = u \quad \forall u, v \in V; \frac{p}{q}, \frac{r}{s} \in \mathbf{Q}$.

Further for $\frac{p}{q}, \frac{r}{s} \in \mathbf{Q}, u \in U$. Consider $\frac{p}{q} \left(\frac{r}{s} u \right) = \frac{p}{q} (rv)$
where $u = sv$
= pw where $rv = qw$

Again $\left(\frac{p}{q} \cdot \frac{r}{s}\right)u = \left(\frac{pr}{qs}\right)u = prt$ where $u = qst$.

Now $u = qst \Rightarrow sv = sqt \Rightarrow s(v - qt) = 0 \Rightarrow v - qt = 0$ as $s \neq 0$ and V has no non-zero element of finite order. Also because of similar reason $qw = rv \Rightarrow qw = qrt \Rightarrow w = rt$.

So we get $\left(\frac{p}{q} \cdot \frac{r}{s}\right)u = prt = pw$.

Thus $\left(\frac{p}{q} \cdot \frac{r}{s}\right)u = \frac{p}{q} \left(\frac{r}{s} u \right)$.

Hence V is a vector space over $\mathbf{Q}$.

Conversely let V be a vector space over $\mathbf{Q}$.

Observe that if V is a vector space over a field F , then for any $n \in \mathbf{Z}$, $\lambda \in F$, $v \in V$, $n(\lambda v) = (n\lambda)v = \lambda(nv)$. (Justify!)

Let now $u \in V$ such that $nu = 0$. If $n \neq 0$ then $\frac{1}{n} \in \mathbf{Q}$ and as V is a vector space over $\mathbf{Q}$ we get that $\frac{1}{n}(nu) = 0 \Rightarrow \left(n \frac{1}{n}\right)u = 0 \Rightarrow u = 0$.

Hence V has no element of finite order except 0. Further given any $u \in V$ and $n \neq 0 \in \mathbf{Z}$, $\frac{1}{n} \in \mathbf{Q} \Rightarrow \frac{1}{n}u \in V$. Put $\frac{1}{n}u = v$.

$$\text{Then } nv = n\left(\frac{1}{n}u\right) = \left(n \frac{1}{n}\right)u = u.$$

Hence V is divisible.

PROBLEMS

1. Show that the following are vector spaces over a field K .

(i) The set of all $m \times n$ matrices with elements in K , under the matrix addition and usual multiplication of a matrix by an element of K .

(ii) The set K^∞ of all infinite sequences $x = (x_1, x_2, x_3, \dots, \dots)$ where $x_i \in K \forall i = 1, 2, \dots$.

Addition and multiplication by scalars of these sequences are defined as under.

For all $x = (x_1, x_2, x_3, \dots, \dots)$, $y = (y_1, y_2, y_3, \dots, \dots)$, $x + y = (x_1 + y_1, x_2 + y_2, x_3 + y_3, \dots, \dots)$ and for all $\lambda \in K$, $\lambda x = (\lambda x_1, \lambda x_2, \lambda x_3, \dots, \dots)$.

2. Show that the set of all polynomials in $K[x]$ of degree at most n (where n is a fixed positive integer) is a subspace of the vector space of Example 2.

3. Show that the set of all infinite sequences $(x_1, x_2, \dots, \dots)$ such that $\sum x_i$ converges is a subspace of $\mathbf{R}^\infty$.

4. Show that union of two subspaces W_1 and W_2 of a vector space V is a subspace of V if and only if either $W_1 \subseteq W_2$ or $W_2 \subseteq W_1$.

5. Prove that if F is a field having more than k elements then V_F cannot be expressed as the set-theoretic union of k proper subspaces.

6. If L, M, N are three subspaces of a vector space V such that $M \subseteq L$. Then show that

$$L \cap (M + N) = M + (L \cap N).$$

Give an example to show that if L, M, N are subspaces of V then $L \cap (M + N) = L \cap M + L \cap N$ is not necessarily true.

7. Prove that if L, M, N are subspaces of a vector space V then

$$L + [M \cap (L + N)] = (L + M) \cap (L + N).$$

8. Do the subspaces of a vector space V form an additive subgroup under the operation '+'? [Ans. No]

9. Show that if in a vector space V

$$L \cap (M + N) = L \cap M + L \cap N$$

holds for all subspaces, L, M, N of V then

$$L + (M \cap N) = (L + M) \cap (L + N) \text{ also holds.}$$

10. Consider the subsets V of $\mathbf{C}^3$ consisting of those vectors (z_1, z_2, z_3) for which

(a) z_1 is real.

(b) $z_1 = 0$.

(c) $z_1 + z_2 = 0$.

(d) $z_1 + z_2 = 1$.

In which of these cases, is V a subspace of $\mathbf{C}^3$? [Ans. (b) and (c)]

11. Let $P = \mathbf{R}[x]$. Consider the subsets V of P consisting of those polynomials $f(x)$ for which

(i) $\deg f(x) = 3$.

(ii) $2f(0) = f(1)$.

(iii) $f(t) \geq 0$ for all $t \in [0, 1]$

(iv) $f(t) = f(1-t)$ for all $t \in \mathbf{R}$.

In which of these cases, is V a subspace of $P_{\mathbf{R}}$? [Ans. (ii) and (iv)]

12. If V is a vector space over a field F , prove that for any integer $n, \lambda \in F, v \in V, n(\lambda v) = \lambda(nv) = (n\lambda)v$. [Ans. (ii) and (iv)]

13. Prove that a non-empty subset W of a vector space V_F is a subspace if and only if for all $\lambda \in F, u, v \in W, \lambda u + v \in W$.

14. Which of the following sets are subspaces of $M_n(F)$, the vector space of $n \times n$ matrices over a field F ?

(a) All upper triangular matrices of order n .

(b) All non-singular matrices of order n .

(c) All symmetric matrices of order n .

(d) All matrices of order n , the sum of whose diagonal entries is zero.

15. A non-empty subset V of the vector space $W_{\mathbf{R}}$ is said to be convex if for all $u, v \in V, cu + dv \in V$ where $c \geq 0, d \geq 0$ and $c + d = 1$. Show that the intersection of arbitrary collection of convex subsets is a convex subset. [Ans. (a), (c) and (d)]

16. Show that in the definition of a vector space V over a field F , the condition $1.x = x \forall x \in V$ can be replaced by the condition ' $\lambda x = 0$ holds only if $\lambda = 0$ or $x = 0, \lambda \in F, x \in V$ '.

17. Show that the set of all solutions of the differential equation

$$\frac{d^2y}{dt^2} + p \frac{dy}{dt} + qy = 0$$

where p and q are fixed functions of t , is a vector space.

18. Let V be the vector space of all functions from $\mathbf{R}$ to $\mathbf{R}$ (cf Example 3); let V_e be the subset of even functions, $f(-x) = f(x)$ and let V_o be the subset of odd functions, $f(-x) = -f(x)$.

Prove: (a) V_e and V_o are subspaces of V .

(b) $V_e + V_o = V$.

(c) $V_e \cap V_o = \{0\}$.

3. Linear Dependence

For any subset X of a vector space V_F , we defined in the last section the concept of the subspace $\langle X \rangle$ spanned by X . When X is non-void and finite, we proved that $\langle X \rangle$ consists of all linear combination of members of X . In fact for any subset $X (\neq \emptyset)$ of V_F , we can show that $\langle X \rangle$ consists of all linear combinations of members of X (taken finitely many at a time). Thus $x \in \langle X \rangle$ if and only if x is a linear combination of finite number of members of X . In this case we say that x is *linearly dependent* upon X . Also we have the following.

Definition 11.13. Let $x_1, x_2, \dots, x_n$ be a finite number of members (not necessarily all distinct) of a vector space V_F . These vectors are said to be *linearly dependent* (L.D.) if for some $\alpha_1, \alpha_2, \dots, \alpha_n \in F$ with at least one of them non-zero, $\alpha_1 x_1 + \alpha_2 x_2 + \dots + \alpha_n x_n = 0$. These vectors are said to be *linearly independent* (L.I.) if for all $\alpha_i \in F$ ($i=1, 2, \dots, n$), $\alpha_1 x_1 + \alpha_2 x_2 + \dots + \alpha_n x_n = 0 \Rightarrow \alpha_i = 0 \forall i$.

A set $X = \{x_1, x_2, \dots, x_n\}$ having n distinct members is said to be L.D. or L.I. according as its n members are L.D. or L.I. In general any subset Y of V_F is said to be L.I. if every finite non-void subset of Y is L.I., otherwise it is called L.D. Thus if $x (\neq 0) \in V$ then as for any $\alpha \in F$, $\alpha x = 0 \Rightarrow \alpha = 0$, we can say that x is linearly independent or in other words the singleton $\{x\}$ is L.I. A null set $\emptyset$ of vectors is L.I., since it has no non-void finite subset and consequently it satisfies the condition defining linearly independent set vacuously. Since for every $\alpha (\neq 0) \in F$, $\alpha 0 = 0$, we see that 0 is L.D. Thus any set of vectors to which 0 belongs, is always L.D.

EXAMPLE 11. Consider $V = \{(\alpha, \beta) \mid \alpha, \beta \in F\}$ where F is a field. The vectors $e_1 = (1, 0)$, $e_2 = (0, 1)$ in V are L.I., since for any $\alpha_1, \alpha_2 \in F$, $\alpha_1 e_1 + \alpha_2 e_2 = 0 \Rightarrow (\alpha_1, 0) + (0, \alpha_2) = (0, 0) \Rightarrow (\alpha_1, \alpha_2) = (0, 0) \Rightarrow \alpha_1 = 0, \alpha_2 = 0$. Further notice that any vector $(\alpha, \beta) \in V$ can be expressed as $\alpha e_1 + \beta e_2$ i.e. as a linear combination of e_1 and e_2 .

Theorem 11.14. If $v_1, v_2, \dots, v_n$ are L.I. in a vector space V_F , then each element of the subspace W spanned by them is expressible uniquely as a linear combination of $v_1, v_2, \dots, v_n$.

Proof: Now each member $w \in W$ is expressible as a linear combination of the given vectors (Theorem 11.9). Suppose some $w \in W$ has two expressions, say

$$w = \alpha_1 v_1 + \alpha_2 v_2 + \dots + \alpha_n v_n$$

and
$$w = \beta_1 v_1 + \beta_2 v_2 + \dots + \beta_n v_n.$$

This yields $0 = (\alpha_1 - \beta_1)v_1 + (\alpha_2 - \beta_2)v_2 + \dots + (\alpha_n - \beta_n)v_n$. Since the v_i 's are L.I., the last equation gives $\alpha_i - \beta_i = 0 \forall i = 1, 2, \dots, n$, i.e. $\alpha_i = \beta_i \forall 1 \leq i \leq n$. This proves that $w = \alpha_1 v_1 + \alpha_2 v_2 + \dots + \alpha_n v_n$ is the unique expression of w as a linear combination of v_i 's ■

Theorem 11.15. Let $u_1, u_2, \dots, u_n$ be any n L.I. vectors in a vector space V_F . Any $n+1$ vectors $v_1, v_2, \dots, v_{n+1}$, each of which is a linear combination of $u_1, u_2, \dots, u_n$, are L.D.

Proof: We prove the result by induction on n . If any of the v_i 's is zero, then trivially given $n+1$ vectors are L.D. So we suppose that none of the v_i 's is zero. Now as v_1 and v_2 are both linear combinations of $u_1, u_2, \dots, u_n$ with $n=1$ we get $v_1 = \alpha_1 u_1$, $v_2 = \alpha_2 u_1$ with $\alpha_1 \neq 0 \neq \alpha_2$. This gives $v_1 - \alpha_1 \alpha_2^{-1} v_2 = 0$. Hence v_1 and v_2 are L.D.

To apply induction suppose that the result holds for any $k (< n)$ L.I. vectors. We can now write

$$v_1 = \alpha_{11} u_1 + \alpha_{12} u_2 + \dots + \alpha_{1n} u_n$$

$$v_2 = \alpha_{21} u_1 + \alpha_{22} u_2 + \dots + \alpha_{2n} u_n$$

$$\dots \dots \dots$$

$$v_{n+1} = \alpha_{n+1,1} u_1 + \alpha_{n+1,2} u_2 + \dots + \alpha_{n+1,n} u_n$$

for some scalars α_{ij} in F . If $\alpha_{in} = 0 \forall i = 1, 2, \dots, n+1$ then each of v_i is a linear combination of $n-1$ vectors $u_1, u_2, \dots, u_{n-1}$ and then the induction hypothesis will yield that $v_1, v_2, \dots, v_n$ are L.D. and consequently $v_1, v_2, \dots, v_{n+1}$ are also L.D. So we suppose that at least one of $\alpha_{in} \neq 0$. To be definite let $\alpha_{1n} \neq 0$. Then for each $i = 2, 3, \dots, n+1$, $v_i - \alpha_{in} \alpha_{1n}^{-1} v_1 = (\alpha_{i1} - \alpha_{in} \alpha_{11} \alpha_{1n}^{-1}) u_1 + (\alpha_{i2} - \alpha_{in} \alpha_{12} \alpha_{1n}^{-1}) u_2 + \dots + (\alpha_{i, n-1} - \alpha_{in} \alpha_{1, n-1} \alpha_{1n}^{-1}) u_{n-1}$.

So by the induction hypothesis the vectors $w_i = v_i - \alpha_{in} \alpha_{1n}^{-1} v_1$ ($2 \leq i \leq n+1$) are L.D. Consequently there exist $\beta_2, \beta_3, \dots, \beta_{n+1}$ in F , not all zero, such that $\beta_2 w_2 + \beta_3 w_3 + \dots + \beta_{n+1} w_{n+1} = 0$, i.e. $\beta_2 (v_2 - \alpha_{2n} \alpha_{1n}^{-1} v_1) + \beta_3 (v_3 - \alpha_{3n} \alpha_{1n}^{-1} v_1) + \dots + \beta_{n+1} (v_{n+1} - \alpha_{n+1,n} \alpha_{1n}^{-1} v_1) = 0$

$$\text{i.e. } (-\beta_2 \alpha_{2n} \alpha_{1n}^{-1} - \beta_3 \alpha_{3n} \alpha_{1n}^{-1} - \dots - \beta_{n+1} \alpha_{n+1,n} \alpha_{1n}^{-1}) v_1 + \beta_2 v_2 + \dots + \beta_{n+1} v_{n+1} = 0$$

This gives that $v_1, v_2, \dots, v_{n+1}$ are L.D. Hence by the principle of Mathematical Induction, result holds for all n . ■

Corollary 11.16. If $\{v_1, v_2, \dots, v_n\}$ is a L.I. subset of a vector space V_F , then any subset W of V having more than n vectors each of which can be expressed as a linear combination of $v_1, v_2, \dots, v_n$ must be L.D.

Proof: Since W has more than n members we can choose $n+1$ members say $u_1, u_2, \dots, u_{n+1}$ of W . Above theorem gives that these vectors are L.D. Thus W has a L.D. subset. Consequently W is L.D. ■

Definition 11.17. A subset B of a vector space V_F is called a *basis* of V if

(1) B is linearly independent; and (2) B spans V .

Definition 11.18. A vector space V_F is said to be *finitely generated* if it has a finite subset which spans V itself.

EXAMPLE 12. For any field F , consider $F^{(n)}$, the vector space consisting of all n -tuples of the type $(\alpha_1, \alpha_2, \dots, \alpha_n)$, over F . Let $e_1 = (1, 0, 0, \dots, 0)$, $e_2 = (0, 1, 0, \dots, 0)$, $\dots$, in general

$$e_i = (0, 0, \dots, 0, \underbrace{1}_{\text{ith place}}, 0, \dots, 0). \quad (i=1, 2, \dots, n).$$

Then as seen in Example 9, $e_1, e_2, \dots, e_n$ are L.I. and they span $F^{(n)}$. Thus $\{e_1, e_2, \dots, e_n\}$ is a basis of $F^{(n)}$ and $F^{(n)}$ is finitely generated.

Lemma 11.19. If $x_1, x_2, \dots, x_n$ are L.D. in a vector space V_F then for some i , $1 \leq i \leq n$, x_i is a linear combination of $x_1, x_2, \dots, x_{i-1}$. In that case the subspace spanned by $\{x_1, x_2, \dots, x_n\}$ is same as the subspace spanned by $\{x_1, x_2, \dots, x_{i-1}, x_{i+1}, \dots, x_n\}$.

Proof: Since $x_1, x_2, \dots, x_n$ are L.D., there exist $\alpha_1, \alpha_2, \dots, \alpha_n$ in F , not all zero such that

$$\alpha_1 x_1 + \alpha_2 x_2 + \dots + \alpha_n x_n = 0. \quad \dots (1)$$

Let i be the largest suffix such that $\alpha_i \neq 0$. Then $\alpha_{i+1}, \alpha_{i+2}, \dots, \alpha_n$ are all zero.

Consequently (1) gives

$$\alpha_1 x_1 + \alpha_2 x_2 + \dots + \alpha_i x_i = 0 \quad \dots (2)$$

This implies that $x_i = -\alpha_i^{-1} \alpha_1 x_1 - \alpha_i^{-1} \alpha_2 x_2 - \dots - \alpha_i^{-1} \alpha_{i-1} x_{i-1}$
 $= \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_{i-1} x_{i-1}$

where $\beta_j = -\alpha_i^{-1} \alpha_j$ for $1 \leq j \leq i-1$. $\dots (3)$

This shows that x_i is a linear combination of $x_1, x_2, \dots, x_{i-1}$.

Let W be the subspace of V spanned by $x_1, x_2, \dots, x_n$. For any $x \in W$, $x = \gamma_1 x_1 + \gamma_2 x_2 + \dots + \gamma_{i-1} x_{i-1} + \gamma_i x_i + \gamma_{i+1} x_{i+1} + \dots + \gamma_n x_n$
 for some $\gamma_j \in F$; $1 \leq j \leq n$.

By using (3) we get

$$\begin{aligned}
 x &= \gamma_1 x_1 + \gamma_2 x_1 + \dots + \gamma_{i-1} x_{i-1} + \gamma_i (\beta_1 x_1 + \beta_2 x_2 \dots + \beta_{i-1} x_{i-1}) \\
 &\quad + \gamma_{i+1} x_{i+1} + \dots + \gamma_n x_n \\
 &= (\gamma_1 + \gamma_i \beta_1) x_1 + (\gamma_2 + \gamma_i \beta_2) x_2 + \dots + (\gamma_{i-1} + \gamma_i \beta_{i-1}) x_{i-1} \\
 &\quad + \gamma_{i+1} x_{i+1} + \dots + \gamma_n x_n.
 \end{aligned}$$

This shows that every $x \in W$ is a linear combination of $x_1, x_2, \dots, x_{i-1}, x_{i+1}, \dots, x_n$. Hence these $n-1$ vectors span W . ■

Theorem 11.20. Let V_F be a finitely generated vector space. Then V_F has a finite basis and any two bases of V_F have some number of vectors.

Proof: Since V_F is finitely generated it has a finite subset $B = \{x_1, x_2, \dots, x_m\}$ which spans V . If B is L.I. then B is desired finite basis of V , otherwise for some $i, 1 \leq i \leq n$, $B_1 = \{x_1, x_2, \dots, x_{i-1}, x_{i+1}, \dots, x_m\}$ also spans V (Lemma 11.19). Then B_1 is a subset of B having $m-1$ vectors which also spans V . Again if B_1 is L.I. we are finished otherwise, as before, B_1 has a subset B_2 containing $m-2$ vectors, which also spans V . Since the number of vectors in B is only m , the process cannot be carried on, for more than m times. Thus after finite number of steps we shall get a subset B_j of B which spans V and is L.I. Then B_j is a desired finite basis of V .

Suppose B_j has n members. Let B' be another basis of V and let B' have more than n members. Since every member of V and in particular every member of B' is a linear combination of the n members of B_j , B' is L.D. (Corollary 11.16). This is a contradiction. Hence the number, say n' , of members of B' cannot exceed n . Interchanging the roles of B' and B_j we also get $n \geq n'$. Hence $n = n'$. This proves that every basis of V has n members. Hence the theorem follows. ■

The above theorem leads us to define the following:

Definition 11.21. If a vector space V_F has a finite basis having n members, then n is called the *dimension* of V . If V does not have a finite basis then V is said to be an *infinite-dimensional vector space*.

As seen in Theorem 11.20 if V_F is finitely generated, then all bases of V_F have same number, say n , of vectors. So the dimension of V_F is uniquely defined. If $V_F = (0)$, then the null set ϕ spans V_F , so that V_F has dimension zero.

REMARK. We are only interested in studying finite-dimensional vector spaces. Hence in this chapter every vector space considered will be finite-dimensional unless otherwise mentioned.

NOTATION. $\dim V$ will denote the dimension of a vector space V .

Theorem 11.22. Let V_F be a vector space of dimension n , then

- (1) any $n+1$ members of V_F are L.D. and
- (2) if $z_1, z_2, \dots, z_n$ span V then they are L.I.

Proof: Since $\dim V = n$, V has a basis $\{x_1, x_2, \dots, x_n\}$. Let $y_1, y_2, \dots, y_{n+1}$ be $n+1$ members of V . By the definition of a basis each of these y_i 's is a linear combination of $x_1, x_2, \dots, x_n$. Hence by Theorem 11.15, $y_1, y_2, \dots, y_{n+1}$ are L.D. This proves (1).

We saw in the proof of Theorem 11.20 that if a finite subset B of V spans V then B has a subset B' which is a basis of V . Since $B = \{z_1, z_2, \dots, z_n\}$ spans V , it has a subset B' which is a basis of V . As $\dim V = n$, B' must have n members. So that $B = B'$ and B itself is a basis of V . Hence $z_1, z_2, \dots, z_n$ are L.I. This proves (2) ■

Theorem 11.23. *If $\{u_1, u_2, \dots, u_k\}$ is a L.I. subset of a finite-dimensional vector space V_F , then it can be extended to a basis of V ; i.e. it is possible to find vectors $v_{k+1}, v_{k+2}, \dots, v_n$ in V such that $\{u_1, u_2, \dots, u_k, v_{k+1}, v_{k+2}, \dots, v_n\}$ is a basis of V .*

Proof: Let $\dim V = n$. Then any $n+1$ vectors in V are L.D. (Theorem 11.22). So $k \leq n$.

Let $\{w_1, w_2, \dots, w_n\}$ be a basis of V . We prove the result by induction on k . For $k=1$, since $u_1 \in V$, we have $u_1 = \alpha_1 w_1 + \alpha_2 w_2 + \dots + \alpha_n w_n$ for some $\alpha_i \in F$. As $u_1 \neq 0$; for some i , $\alpha_i \neq 0$. If j is the largest suffix such that $\alpha_j \neq 0$ then

$$\begin{aligned} u_1 &= \alpha_1 w_1 + \alpha_2 w_2 + \dots + \alpha_j w_j \\ \Rightarrow w_j &= \alpha_j^{-1} u_1 - \alpha_j^{-1} \alpha_1 w_1 - \dots - \alpha_j^{-1} \alpha_{j-1} w_{j-1} \\ \Rightarrow w_j &\in \langle u_1, w_1, w_2, \dots, w_{j-1} \rangle \\ &\subseteq \langle u_1, w_1, w_2, \dots, w_{j-1}, w_{j+1}, \dots, w_n \rangle. \end{aligned}$$

Since for every $i \neq j$, we also have

$$w_i \in \langle u_1, w_1, w_2, \dots, w_{j-1}, w_{j+1}, \dots, w_n \rangle, \text{ we get} \\ V = \langle w_1, w_2, \dots, w_n \rangle \subseteq \langle u_1, w_1, w_2, \dots, w_{j-1}, w_{j+1}, \dots, w_n \rangle \subseteq V.$$

$$\text{Hence } V = \langle u_1, w_1, w_2, \dots, w_{j-1}, w_{j+1}, \dots, w_n \rangle.$$

This shows that n vectors $u_1, w_1, w_2, \dots, w_{j-1}, w_{j+1}, \dots, w_n$ generate V . As $\dim V = n$, these n vectors form a basis of V . So the result holds for $k=1$. To apply induction let $k > 1$ and let the result holds for $k-1$ L.I. vectors. Then we can find vectors $t_k, t_{k+1}, \dots, t_n$ such that $\{u_1, u_2, \dots, u_{k-1}, t_k, t_{k+1}, \dots, t_n\}$ is a basis of V . As $u_k \in V$ we have $u_k = \beta_1 u_1 + \beta_2 u_2 + \dots + \beta_{k-1} u_{k-1} + \gamma_k t_k + \dots + \gamma_n t_n$ for some $\beta_i, \gamma_i \in F$.

If $\gamma_k = \gamma_{k+1} = \dots = \gamma_n = 0$ then we get $u_k = \beta_1 u_1 + \beta_2 u_2 + \dots + \beta_{k-1} u_{k-1} \Rightarrow u_1, u_2, \dots, u_{k-1}, u_k$ are L.D. This is a contradiction.

So for some $i \geq k$, $\gamma_i \neq 0$. Then

$$\begin{aligned} t_i &= -\gamma_i^{-1} \beta_1 u_1 - \dots - \gamma_i^{-1} u_k - \gamma_i^{-1} \gamma_k t_k - \dots - \gamma_i^{-1} \gamma_{i-1} t_{i-1} \\ &\quad - \gamma_i^{-1} \gamma_{i+1} t_{i+1} - \dots - \gamma_i^{-1} \gamma_n t_n; \end{aligned}$$

hence $t_i \in \langle u_1, u_2, \dots, u_k, t_k, t_{k+1}, \dots, t_{i-1}, t_{i+1}, \dots, t_n \rangle$. As before, this gives that $\langle u_1, u_2, \dots, u_k, t_k, t_{k+1}, \dots, t_{i-1}, t_{i+1}, \dots, t_n \rangle$ is a basis of V . Hence by induction the result holds for all $k \leq n$. ■

Corollary 11.24. *If a vector space V_F is of dimension n , then any n L.I. vectors in V_F form a basis of V_F .*

Proof: Let $x_1, x_2, \dots, x_n$ be any n L.I. vectors in V . The above theorem gives that these n vectors can be adjoined with some vectors to form a basis of V . However any basis of V contains only n members. Consequently these n vectors form a basis of V . ■

Corollary 11.25. *For any subspace W of a finite-dimensional vector space V_F , $\dim W \leq \dim V$. Further $W=V$ if and only if $\dim W = \dim V$.*

Proof: Let $\dim V = n$. Since V cannot contain more than n L.I. vectors, W cannot have any L.I. subset containing more than n members.

So we can find a L.I. subset $B = \{y_1, y_2, \dots, y_m\}$ of W containing maximum number, say m , of members. Then $m \leq n$ and any $m+1$ vectors in W are L.D. Thus if $y \neq 0 \in W$, then $y_1, y_2, \dots, y_m, y$ are L.D. and so there exist $\beta_1, \beta_2, \dots, \beta_m, \beta$ in F not all zero such that $\beta_1 y_1 + \beta_2 y_2 + \dots + \beta_m y_m + \beta y = 0$. Here $\beta \neq 0$ since otherwise the linear independence of $y_1, y_2, y_3, \dots, y_m$ will force $\beta_1, \beta_2, \dots, \beta_m$ all to be zero. So we have

$$y = -\beta^{-1}\beta_1 y_1 - \beta^{-1}\beta_2 y_2 - \dots - \beta^{-1}\beta_m y_m.$$

This gives that every member of W is expressible as a linear combination of $y_1, y_2, \dots, y_m$. Consequently $\{y_1, y_2, \dots, y_m\}$ is a basis of W . This proves the first part.

Now if W has a basis $\{z_1, z_2, \dots, z_n\}$ of n members, these n members are also L.I. in V . Since any n L.I. members of V form a basis of V ; we get $\{z_1, z_2, \dots, z_n\}$ spans V . So we have $W=V$. Further $W=V$ clearly implies that $\dim W = \dim V$. ■

Theorem 11.26. *If U and W are subspaces of a finite-dimensional vector space V_F , then $\dim(U+W) = \dim U + \dim W - \dim(U \cap W)$.*

Proof: Let $\dim U = r$, $\dim W = s$ and $\dim(U \cap W) = t$.

Let $\{e_1, e_2, \dots, e_t\}$ be a basis of $U \cap W$. Since $U \cap W$ is a subspace of U we can extend the above basis to a basis $\{e_1, e_2, \dots, e_t, f_1, f_2, \dots, f_{r-t}\}$ of U . This basis has r members, since $\dim U = r$. Similarly we can find a basis $\{e_1, e_2, \dots, e_t, g_1, g_2, \dots, g_{s-t}\}$ of W . Now we claim that the set

$$B = \{e_1, e_2, \dots, e_t, f_1, f_2, \dots, f_{r-t}, g_1, g_2, \dots, g_{s-t}\}$$

containing $r+s-t$ vectors is a basis of $U+W$ and consequently $\dim(U+W) = r+s-t$.

To see this consider any $x \in U+W$. Then $x = u + v$ for some $u \in U$, $v \in W$.

Further $u = \alpha_1 e_1 + \alpha_2 e_2 + \dots + \alpha_t e_t + \beta_1 f_1 + \beta_2 f_2 + \dots + \beta_{r-t} f_{r-t}$
 and $v = \alpha'_1 e_1 + \alpha'_2 e_2 + \dots + \alpha'_t e_t + \gamma_1 g_1 + \gamma_2 g_2 + \dots + \gamma_{s-t} g_{s-t}$
 for some $\alpha_i, \alpha'_j, \beta_k, \gamma_l \in F$.

This gives $x = u + v = (\alpha_1 + \alpha'_1) e_1 + (\alpha_2 + \alpha'_2) e_2 + \dots$
 $+ (\alpha_t + \alpha'_t) e_t + \beta_1 f_1 + \beta_2 f_2 + \dots + \beta_{r-t} f_{r-t} + \gamma_1 g_1 + \gamma_2 g_2 + \dots$
 $+ \gamma_{s-t} g_{s-t}.$

Thus every vector x in $U+W$ is a linear combination of vectors in B . So B spans $U+W$. To show that B is also L.I., let

$$\alpha_1 e_1 + \alpha_2 e_2 + \dots + \alpha_t e_t + \beta_1 f_1 + \beta_2 f_2 + \dots + \beta_{r-t} f_{r-t} \\ + \gamma_1 g_1 + \gamma_2 g_2 + \dots + \gamma_{s-t} g_{s-t} = 0 \quad \dots (1)$$

for some $\alpha_i, \beta_j, \gamma_k$ in F .

This gives $-(\alpha_1 e_1 + \alpha_2 e_2 + \dots + \alpha_t e_t + \beta_1 f_1 + \beta_2 f_2 + \dots$
 $\dots + \beta_{r-t} f_{r-t}) = \gamma_1 g_1 + \gamma_2 g_2 + \dots + \gamma_{s-t} g_{s-t} \in U \cap W$.

But $U \cap W = \langle e_1, e_2, \dots, e_t \rangle$.

$\Rightarrow \gamma_1 g_1 + \gamma_2 g_2 + \dots + \gamma_{s-t} g_{s-t} = \delta_1 e_1 + \delta_2 e_2 + \dots + \delta_t e_t$ for some $\delta_i \in F$.

$\Rightarrow \delta_1 e_1 + \delta_2 e_2 + \dots + \delta_t e_t - \gamma_1 g_1 - \gamma_2 g_2 - \dots - \gamma_{s-t} g_{s-t} = 0$.

$\Rightarrow \gamma_1 = \gamma_2 = \gamma_3 = \dots = \gamma_{s-t} = 0$ since $e_1, e_2, \dots, e_t, g_1, g_2, \dots, g_{s-t}$
 are L.I. Then (1) yields, $\alpha_1 e_1 + \alpha_2 e_2 + \dots + \alpha_t e_t + \beta_1 f_1 + \beta_2 f_2 + \dots$
 $\dots + \beta_{r-t} f_{r-t} = 0$

$\Rightarrow \alpha_1 = \alpha_2 = \alpha_3 = \dots = \alpha_t = \beta_1 = \beta_2 = \dots = \beta_{r-t} = 0$, since $e_1, e_2, \dots, e_t,$
 $f_1, f_2, \dots, f_{r-t}$ are L.I.

Consequently all the coefficients of L.H.S. of (1) are zero. This
 proves that B is also L.I. Hence B is a basis of $U+W$ and we get

$$\dim(U+W) = \dim U + \dim W - \dim(U \cap W). \blacksquare$$

Corollary 11.27. If U and W are two subspaces of a finite-dimensional vector space V_F such that $U \cap W = \{0\}$ then $\dim(U+W) = \dim U + \dim W$.

Proof: Since $U \cap W = \{0\}$, $\dim(U \cap W) = 0$.

Thus $\dim(U+W) = \dim U + \dim W - \dim(U \cap W)$ gives that
 $\dim(U+W) = \dim U + \dim W$.

WORKED-OUT EXERCISES

Exercise 1. Show that the vectors (ξ_1, ξ_2) and (η_1, η_2) in $\mathbb{C}^2$ are linearly dependent if and only if $\xi_1 \eta_2 = \xi_2 \eta_1$.

Solution. Let (ξ_1, ξ_2) and (η_1, η_2) be L.D.

There exist $\alpha_1, \alpha_2 \in \mathbb{C}$ such that $\alpha_1(\xi_1, \xi_2) + \alpha_2(\eta_1, \eta_2) = (0, 0)$ with at least one of α_1, α_2 being non-zero. Without loss of generality, let us assume that $\alpha_1 \neq 0$.

This implies $(\xi_1, \xi_2) = \alpha_1^{-1} \alpha_2 (\eta_1, \eta_2)$

$$\Rightarrow \xi_1 = -\alpha_1^{-1} \alpha_2 \eta_1 \quad \text{and} \quad \xi_2 = -\alpha_1^{-1} \alpha_2 \eta_2$$

$$\Rightarrow \xi_1 \eta_2 = -\alpha_1^{-1} \alpha_2 \eta_1 \eta_2 = (-\alpha_1^{-1} \alpha_2 \eta_2) \eta_1 = \xi_2 \eta_1.$$

Conversely let $\xi_1 \eta_2 = \xi_2 \eta_1$ (1)

If $\xi_1 = \xi_2 = 0$, then $(\xi_1, \xi_2) = (0, 0)$ and so (ξ_1, ξ_2) and (η_1, η_2) are L.D.

Let us suppose that $\xi_1 \neq 0$. Then $\frac{\eta_1}{\xi_1} \in \mathbf{C}$.

$$\begin{aligned} \text{Also } -\frac{\eta_1}{\xi_1} (\xi_1, \xi_2) + (\eta_1, \eta_2) &= \left(-\eta_1, \frac{-\eta_1 \xi_2}{\xi_1} \right) + (\eta_1, \eta_2) \\ &= (-\eta_1, -\eta_2) + (\eta_1, \eta_2) = (0, 0) \quad \text{by (1).} \end{aligned}$$

Hence (ξ_1, ξ_2) and (η_1, η_2) are L.D.

Exercise 2. Let $P = \mathbf{R}[x]$ and let $f(x), g(x), h(x), k(x) \in P$ be such that $f(t) = 1, g(t) = t, h(t) = t^2$ and $k(t) = 1 + t + t^2$ for all $t \in \mathbf{R}$. Show that $f(x), g(x), h(x), k(x)$ are linearly dependent but any three of them are linearly independent.

Solution. Consider the element $f(x) + g(x) + h(x) - k(x)$. For all $t \in \mathbf{R}$, $f(t) + g(t) + h(t) - k(t) = 1 + t + t^2 - (1 + t + t^2) = 0$.

So $f(x) + g(x) + h(x) - k(x)$ is zero polynomial (Recall that any polynomial of degree n with real coefficients have at most n roots. We shall prove this result in a more general setting in chapter on Fields).

Hence $f(x), g(x), h(x)$ and $k(x)$ are L.D.

We now prove any three of them are L.I.

We shall consider $f(x), g(x), h(x)$ and $f(x), h(x), k(x)$.

Remaining two triads can be discussed in a similar fashion.

Let $\alpha_1 f(x) + \alpha_2 g(x) + \alpha_3 h(x) = 0$ for some $\alpha_1, \alpha_2, \alpha_3 \in \mathbf{R}$. This gives that $\alpha_1 f(t) + \alpha_2 g(t) + \alpha_3 h(t) = 0 \quad \forall t \in \mathbf{R}$

$$\Rightarrow \alpha_1 + \alpha_2 t + \alpha_3 t^2 = 0 \quad \forall t \in \mathbf{R}$$

$$\Rightarrow \alpha_1 = \alpha_2 = \alpha_3 = 0.$$

So $f(x), g(x), h(x)$ are L.I.

Finally $\beta_1 f(x) + \beta_2 h(x) + \beta_3 k(x) = 0$

$$\Rightarrow \beta_1 + \beta_2 t^2 + \beta_3 (1 + t + t^2) = 0 \quad \forall t \in \mathbf{R}.$$

Thus we get $\beta_1 + \beta_3 = 0, \beta_3 = 0, \beta_2 + \beta_3 = 0$.

Therefore, $\beta_1 = \beta_2 = \beta_3 = 0$.

Hence $f(x), h(x)$ and $k(x)$ are L.I.

Exercise 3. Under what condition, on the scalar ξ do the vectors $(1, 1, 1)$ and $(1, \xi, \xi^2)$ form a basis of $\mathbf{C}^3$?

Solution. Since any basis of $\mathbf{C}^3$ must consist of three vectors, ξ must take two values such that ξ^2 is same. In another word we are to find condition on scalar x such that $(1, 1, 1), (1, x, x^2)$ and $(1, -x, x^2)$ form a basis of $\mathbf{C}^3$. As $\dim \mathbf{C}^3$ over $\mathbf{C}$ is 3, it is sufficient to find the condition such that $(1, 1, 1), (1, x, x^2)$ and $(1, -x, x^2)$ are L.I. over $\mathbf{C}$.

These vectors are L.I. over $\mathbf{C}$ if and only

$$\lambda_1(1, 1, 1) + \lambda_2(1, x, x^2) + \lambda_3(1, -x, x^2) = (0, 0, 0) \\ \Rightarrow \lambda_1 = \lambda_2 = \lambda_3 = 0 \vee \lambda_1, \lambda_2, \lambda_3 \in \mathbf{C}.$$

Now $\lambda_1(1, 1, 1) + \lambda_2(1, x, x^2) + \lambda_3(1, -x, x^2) = (0, 0, 0)$

$$\Rightarrow \left. \begin{aligned} \lambda_1 + \lambda_2 + \lambda_3 &= 0 \\ \lambda_1 + \lambda_2 x - \lambda_3 x &= 0 \\ \lambda_1 + \lambda_2 x^2 + \lambda_3 x^2 &= 0 \end{aligned} \right\} \dots (1)$$

So the given vectors are L.I. over $\mathbf{C}$ if and only if the system (1) of linear equations has trivial solution.

Now (1) has trivial solution if and only if $\begin{vmatrix} 1 & 1 & 1 \\ 1 & x & -x \\ 1 & x^2 & x^2 \end{vmatrix} \neq 0$

$$2x(x^2 - 1) \neq 0.$$

So the vectors $(1, 1, 1), (1, x, x^2), (1, -x, x^2)$, form a basis of $\mathbf{C}^3$ if and only if $x \notin \{0, 1, -1\}$.

Hence $(1, 1, 1), (1, \xi, \xi^2)$ form a basis of $\mathbf{C}^3$ if and only if $\xi = \pm x$ where $x \notin \{0, 1, -1\}$.

Exercise 4. Suppose that x and y are vectors and U is a subspace of a vector space V_F . Let W be the subspace spanned by U and x and let Y be the subspace spanned by U and y . Prove that if $y \in W - U$ then $x \in Y$.

Solution. Since $y \in W = \langle U, x \rangle$

$$y = \alpha_1 u_1 + \dots + \alpha_i u_i + \beta x \text{ for some}$$

$$u_1, u_2, \dots, u_i \in U \text{ and } \alpha_1, \dots, \alpha_i, \beta \in F.$$

As $y \notin U$; $\beta \neq 0$. Hence $\beta^{-1} \in F$

$$\begin{aligned} x &= \beta^{-1} y - \beta^{-1} \alpha_1 u_1 - \dots - \beta^{-1} \alpha_i u_i \\ &= \beta^{-1} y + (-\beta^{-1} \alpha_1) u_1 + \dots + (-\beta^{-1} \alpha_i) u_i \\ &\in \langle U, y \rangle = Y. \end{aligned}$$

Exercise 5. If x, y and z are vectors in V_F such that $x + y + z = 0$, then x and y span the same subspace as y and z .

Solution. Let $W = \langle x, y \rangle$ and $U = \langle y, z \rangle$.

Let $w \in W$, then $w = \alpha x + \beta y$ for some $\alpha, \beta \in F$

$$\begin{aligned} &= \alpha(-y - z) + \beta y \quad \text{as } x + y + z = 0 \\ &= (\beta - \alpha)y + (-\alpha)z \\ &\in \langle y, z \rangle = U \end{aligned}$$

This gives $W \subseteq U$.

For reverse inclusion, we proceed as under.

Let $u \in U$, then $u = \gamma y + \delta z$ for some $\gamma, \delta \in F$

$$\begin{aligned} &= \gamma y + \delta(-x - y) \quad \text{since } x + y + z = 0 \\ &= (-\delta)x + (\gamma - \delta)y \\ &\in \langle x, y \rangle = W. \end{aligned}$$

This implies that $W \subseteq U$.

Hence $U = W$.

PROBLEMS

1. Prove that the vectors $(1, 1, 0, 0)$, $(0, 1, -1, 0)$ and $(0, 0, 0, 3)$ in $F^{(4)}$ are linearly independent over F , if F is a field of characteristic zero. Show also that the same vectors are linearly dependent over F if F is of characteristic 3.

2. Prove that the four vectors $(1, 0, 0)$, $(0, 1, 0)$, $(0, 0, 1)$ and $(1, 1, 1)$ in $\mathbf{C}^3$ form a linearly dependent set over $\mathbf{C}$ but any three of them are linearly independent over $\mathbf{C}$.

3. Prove that $\mathbf{C}_R$ is a two-dimensional vector space.

4. Show that the vector space of all polynomials in $\mathbf{R}[x]$ over $\mathbf{R}$ is an infinite-dimensional vector space.

5. Let S be the set of all polynomials in $\mathbf{R}[x]$ of degree at most n (n is a fixed positive integer). Show that dimension of S_R is n .

6. Show that in a vector space every subset of a linearly independent set is linearly independent and every superset of a linearly dependent set is linearly dependent.

7. Let V be a vector space over $\mathbf{R}$ and let $T = \{(x, y) \mid x, y \in V\}$. Define addition in T componentwise and scalar multiplication by a complex number $\alpha + \beta i$ by $(\alpha + \beta i)(x, y) = (\alpha x - \beta y, \beta x + \alpha y)$. Show that T is a vector space over $\mathbf{C}$. Show that $\dim T$ over $\mathbf{C} = \dim V$ over $\mathbf{R}$.

8. Let $\{U_1, U_2, \dots, U_l\}$ be any collection of subspaces of an n -dimensional vector space V such that the intersection of any n of these subspaces has dimension at least one. Prove that $\dim(U_1 \cap \dots \cap U_l) \geq 1$.

9. Prove that $F[x]$, is an infinite dimensional vector space over F , where F is a field. (This is generalization of problem 4 above)

10. If $\mathbf{R}$ is considered as a vector space over $\mathbf{Q}$, then show that a necessary and sufficient condition that the vectors 1 and ξ in $\mathbf{R}$ be linearly independent is that the real number ξ be irrational.

11. Show that if x, y, z are linearly independent vectors then so are $x+y, y+z$ and $z+x$ also, provided $\text{char } F \neq 2$.

12. Under what conditions on the scalar ξ do the vectors $(0, 1, \xi)$, $(\xi, 0, 1)$ and $(\xi, 1, 1+\xi)$ form a basis of $\mathbf{C}^3$.

[Ans. Under no condition]

13. If S and T are non-empty subsets of the vector space V , establish that

$$(a) \ S \subseteq T \Rightarrow \langle S \rangle \subseteq \langle T \rangle.$$

$$(b) \ \langle S \cup T \rangle = \langle S \rangle + \langle T \rangle.$$

(c) $S = \langle S \rangle$ if and only if S is subspace of V .

(d) $\langle \langle S \rangle \rangle = \langle S \rangle$.

14. Show that $\{(1, \alpha, \alpha^2), (1, \beta, \beta^2), (1, \gamma, \gamma^2)\}$ is a linearly independent subset of $\mathbf{R}^3$ for all real numbers α, β and γ . When it is a basis? [Ans. The given set is a basis when it has three distinct elements, i.e. $\alpha \neq \beta, \beta \neq \gamma$, and $\alpha \neq \gamma$]

15. If for each natural number n the function f_n is defined by $f_n(x) = e^{nx} \forall x \in \mathbf{R}$, then show that f_1, f_2, f_3 are linearly independent vectors in $\mathbf{R}^{\mathbf{R}}$ (The vector space over $\mathbf{R}$ of all real valued-functions.).

4. Quotient Spaces

Let W be a subspace of a vector space V_F . Since W is a subgroup of the additive group $\langle V, + \rangle$, we can define the quotient group $V/W = \{x+W \mid x \in V\}$ such that for any $x+W, y+W$ in V/W

$$(x+W) + (y+W) = (x+y) + W$$

This group, we know, is Abelian. If for any $\alpha \in F, x+W$ in V/W , we define

$$\alpha(x+W) = \alpha x + W \quad \dots (1)$$

Then we see that $x+W = y+W$

$$\Rightarrow x - y \in W \Rightarrow \alpha(x - y) \in W \Rightarrow \alpha x - \alpha y \in W$$

$$\Rightarrow \alpha x + W = \alpha y + W.$$

This shows that the multiplication of $x+W$ by $\alpha \in F$ is well-defined.

Now for any $\alpha, \beta \in F; x+W, y+W \in V/W$, we see that

$$(\alpha + \beta)(x+W) = (\alpha + \beta)x + W = \alpha x + \beta x + W$$

$$= (\alpha x + W) + (\beta x + W) = \alpha(x+W) + \beta(x+W).$$

$$\alpha[(x+W) + (y+W)] = \alpha[(x+y) + W]$$

$$= \alpha(x+y) + W$$

$$= (\alpha x + \alpha y) + W$$

$$= (\alpha x + W) + (\alpha y + W)$$

$$= \alpha(x+W) + \alpha(y+W).$$

$$(\alpha\beta)(x+W) = (\alpha\beta)x + W$$

$$= \alpha(\beta x) + W = \alpha(\beta x + W)$$

$$= \alpha[\beta(x+W)].$$

$$1(\alpha+W) = 1x+W = x+W$$

This shows that V/W is a vector space over F .

So we define the following:

Definition 11.28. Let W be a subspace of a vector space V_F . Then set $V/W = \{x+W \mid x \in V\}$ is a vector space over F , addition and multiplication of vectors by scalars are defined as

$$(x+W) + (y+W) = (x+y) + W$$

$$\alpha(x+W) = \alpha x + W.$$

for any $x+W, y+W$, in V/W and α in F .

This vector space is called a *Quotient Space*.

Theorem 11.29. Let W be a subspace of a finite-dimensional vector space V_F . Then $\dim V/W = \dim V - \dim W$.

Proof: Let $\dim V = n$, $\dim W = m$. Let $\{x_1, x_2, \dots, x_m\}$ be a basis of W . These m vectors $x_1, x_2, \dots, x_m$ are L.I. in V , so they can be extended to a basis of V , say $\{x_1, x_2, \dots, x_m, x_{m+1}, \dots, x_n\}$. Now consider $n-m$ vectors $x_{m+1}+W, x_{m+2}+W, \dots, x_n+W$. We show that these $n-m$ vectors form a basis of V/W and hence $\dim(V/W) = n-m$.

Now for any $x \in V$, $x = \alpha_1 x_1 + \alpha_2 x_2 + \dots + \alpha_n x_n$ gives

$$\begin{aligned} x+W &= (\alpha_{m+1}x_{m+1} + \dots + \alpha_n x_n) + (\alpha_1 x_1 + \alpha_2 x_2 + \dots + \alpha_m x_m) + W \\ &= \alpha_{m+1}x_{m+1} + \dots + \alpha_n x_n + W, \text{ since } x_i \in W \text{ for every } i=1, 2, \\ &\dots, m \text{ gives } \alpha_1 x_1 + \alpha_2 x_2 + \dots + \alpha_m x_m + W = W. \end{aligned}$$

So $x+W = \alpha_{m+1}(x_{m+1}+W) + \dots + \alpha_n(x_n+W)$.

This shows that the vectors $x_{m+1}+W, x_{m+2}+W, \dots, x_n+W$ span V/W . So it only remains to be shown that these vectors are L.I. Suppose that for some $\beta_{m+1}, \beta_{m+2}, \dots, \beta_n$ in F , $\beta_{m+1}(x_{m+1}+W) + \beta_{m+2}(x_{m+2}+W) + \dots + \beta_n(x_n+W) = \bar{0}$

$$\text{i.e. } (\beta_{m+1}x_{m+1} + \beta_{m+2}x_{m+2} + \dots + \beta_n x_n) + W = W$$

$$\Rightarrow \beta_{m+1}x_{m+1} + \beta_{m+2}x_{m+2} + \dots + \beta_n x_n \in W$$

$$\Rightarrow \beta_{m+1}x_{m+1} + \beta_{m+2}x_{m+2} + \dots + \beta_n x_n = \gamma_1 x_1 + \gamma_2 x_2 + \dots + \gamma_m x_m$$

for some $\gamma_i \in F, 1 \leq i \leq m$.

$$\Rightarrow \gamma_1 x_1 + \gamma_2 x_2 + \dots + \gamma_m x_m + (-\beta_{m+1})x_{m+1} + (-\beta_{m+2})x_{m+2} + \dots + (-\beta_n)x_n = \bar{0}$$

$$\Rightarrow \gamma_1 = \gamma_2 = \dots = \gamma_m = -\beta_{m+1} = -\beta_{m+2} = \dots = -\beta_n = 0$$

since $x_1, x_2, \dots, x_n$ are L.I.

$$\Rightarrow \beta_{m+1}, \beta_{m+2}, \dots, \beta_n \text{ are all zero.}$$

This proves that $x_{m+1}+W, x_{m+2}+W, \dots, x_n+W$ are L.I.

Hence $\dim(V/W) = n-m = \dim V - \dim W$. ■

PROBLEMS

1. If W is a subspace of a vector space V , prove there is 1-1 correspondence between subspaces of V which contain W and subspaces of V/W .

2. Let W be a subspace of a vector space V_F . For $a, b \in W$ we define $a \equiv b \pmod{W}$ [called congruence modulo W] if and only if $a-b \in W$ and for all $\alpha \in F$, $\alpha(a-b) \in W$. Show that congruence modulo W is an equivalence relation on V . Further show that for each $a \in W$, $Cl(a) = a+W$.

3. Give an example of an infinite-dimensional vector space V_F with a subspace W such that V/W is a finite-dimensional vector space.

5. Direct Sums and Complements

Definition 11.30. If U and W are two subspaces of a vector space V such that $U \cap W = (0)$, then the sum $U \oplus W$ is called a *direct sum* of U and W and is denoted by $U \oplus W$.

EXAMPLE 13. Let $V = F[x]$ be the vector space of all polynomials over a field F . Then

$U = \{\alpha_0 + \alpha_1 x \mid \alpha_0, \alpha_1 \in F\}$ and $W = \{\alpha_2 x^2 + \alpha_3 x^3 \mid \alpha_2, \alpha_3 \in F\}$ are two subspaces of V such that $U \cap W = (0)$.

So that $U + W = U \oplus W$.

EXAMPLE 14. Consider $V = F^{(4)}$. Let $U = \{(\alpha, \beta, 0, 0) \mid \alpha, \beta \in F\}$ and $W = \{(0, 0, \gamma, \delta) \mid \gamma, \delta \in F\}$. It is a simple matter to verify that both these subsets are subspaces of V . The zero vector of V is $(0, 0, 0, 0)$. Then $z = (\alpha, \beta, 0, 0) = (0, 0, \gamma, \delta)$ for some $\alpha, \beta, \gamma, \delta \in F \Rightarrow \alpha = \beta = \gamma = \delta = 0 \Rightarrow z = 0$, so that $U \cap W = (0)$.

Hence $U + W = U \oplus W$. Now every $(\alpha_1, \alpha_2, \alpha_3, \alpha_4) \in V$ can be expressed as $(\alpha_1, \alpha_2, 0, 0) + (0, 0, \alpha_3, \alpha_4)$ with $(\alpha_1, \alpha_2, 0, 0) \in U$, $(0, 0, \alpha_3, \alpha_4) \in W$.

This gives $V = U \oplus W$.

Now Corollary 11.27 can be rephrased and put as

COROLLARY 11.27a. If U and W are subspaces of a finite-dimensional vector space V_F such that their sum is direct, then $\dim(U \oplus W) = \dim U + \dim W$.

Definition 11.31. Let U be a subspace of a vector space V_F . Then a subspace U' of V such that $V = U \oplus U'$, is called a *complement* of U .

Thus in Example 14, W is a complement of U .

Theorem 11.32. Every subspace of a finite-dimensional vector space has a complement.

Proof: Let U be a subspace of a finite-dimensional vector space V_F . Further let $\dim V = n$, $\dim U = m$. Since every L.I. subset of V can be extended to a basis of V , we can find a basis $\{x_1, x_2, \dots, x_n\}$ of V such that $\{x_1, x_2, \dots, x_m\}$ is a basis of U . Let U' be the subspace of V spanned by $\{x_{m+1}, x_{m+2}, \dots, x_n\}$. Consider any $x \in V$, then $x = \alpha_1 x_1 + \alpha_2 x_2 + \dots + \alpha_n x_n$ for some $\alpha_i \in F$ i.e. $x = (\alpha_1 x_1 + \alpha_2 x_2 + \dots + \alpha_m x_m) + (\alpha_{m+1} x_{m+1} + \dots + \alpha_n x_n) \in U + U'$. This implies $V = U + U' \dots (1)$

Suppose that $z \in U \cap U'$ then $z = \alpha_1 x_1 + \alpha_2 x_2 + \dots + \alpha_m x_m = \beta_{m+1} x_{m+1} + \dots + \beta_n x_n$ for some α_i & $\beta_j \in F$, since $\{x_1, x_2, \dots, x_m\}$ is a basis of U and $\{x_{m+1}, \dots, x_n\}$ is basis of U' .

Consequently $\alpha_1 x_1 + \alpha_2 x_2 + \dots + \alpha_m x_m + (-\beta_{m+1}) x_{m+1} + (-\beta_{m+2}) x_{m+2} + \dots + (-\beta_n) x_n = 0 \Rightarrow \alpha_1 = \alpha_2 = \dots = \alpha_m = -\beta_{m+1} = -\beta_{m+2} = \dots = -\beta_n = 0 \Rightarrow z = 0$. Thus $U \cap U' = (0)$

Hence (1) yields that $V = U \oplus U'$.

This proves that U' is a complement of U . ■

Corollary 11.33. *If U is a subspace of a finite-dimensional vector space V , then every complement of U has dimension equal to $\dim V - \dim U$.*

Proof: If U' is a complement of U , then $W = U \oplus U'$. So that $\dim V = \dim U + \dim U'$ (Corollary 1.1.27a).

This yields that $\dim U' = \dim V - \dim U$. Hence the corollary follows. ■

Thus the above corollary shows that all complements of a subspace are of same dimension. However the following example shows that a complement is not unique.

EXAMPLE 15. Consider $V = F^{(2)}$. Let $U = \{(\alpha, 0) \mid \alpha \in F\}$

$$U' = \{(0, \beta) \mid \beta \in F\} \text{ and } W = \{(\alpha, \alpha) \mid \alpha \in F\}$$

U, U', W are all subspaces of V . They are spanned by the vectors $(1, 0), (0, 1)$ and $(1, 1)$ respectively. Thus each of them is of dimension one.

Consider $z \in U \cap U' \Rightarrow z = (\alpha, 0) = (0, \beta)$ for some $\alpha, \beta \in F \Rightarrow \alpha = \beta = 0 \Rightarrow z = 0$. Thus $U \cap U' = (0)$. This yields $U + U' = U \oplus U'$. Then $\dim(U + U') = \dim U + \dim U' = 1 + 1 = 2 \Rightarrow V = U \oplus U' \Rightarrow U'$ is complement of U . Again $z \in U \cap W \Rightarrow z = (\alpha, 0) = (\beta, \beta)$ for some $\alpha, \beta \in F \Rightarrow \alpha = \beta = 0 \Rightarrow z = 0 \Rightarrow U \cap W = (0)$. Again we get $V = U \oplus W$. So that W is also a complement of U . Hence U' and W are two distinct complements of U . This shows that complement of a subspace is not unique.

WORKED-OUT EXERCISES

Exercise 1. *Definition: A vector space V_F is said to be direct sum (internal) of its subspaces $W_1, W_2, \dots, W_n$; if each $v \in V$ can be written in one and only one ways as $\sum_{i=1}^n w_i$, with $w_i \in W_i$.*

We write $V = W_1 \oplus W_2 \oplus \dots \oplus W_n$.

Show that if $V = W_1 + W_2 + \dots + W_n$, then following are equivalent

(i) $V = W_1 \oplus W_2 \oplus \dots \oplus W_n$.

(ii) For each $1 \leq i \leq n$, $W_i \cap \sum_{\substack{j=1 \\ j \neq i}}^n W_j = (0)$.

(iii) For each $w_i \in W_i$, $\sum_{i=1}^n w_i = 0 \Rightarrow w_i = 0$.

(iv) If for each $1 \leq i \leq n$, $B_i = \{b_{i1}, \dots, b_{ik_i}\}$ is a basis of W_i , then

$\bigcup_{i=1}^n B_i$ is a basis of V . [Here by $\bigcup_{i=1}^n B_i$, we mean the vectors $b_1^1, b_2^1, b_3^1, \dots, b_{k_1}^1, b_1^2, \dots, b_{k_2}^2, \dots, b_1^n, \dots, b_{k_n}^n$.]

Solution (i) $\Rightarrow$ (ii). Let $x \in W_i \cap \sum_{\substack{j=1 \\ j \neq i}}^n W_j$.

$x = w_1$ and also $x = w_1 + w_2 + \dots + w_{i-1} + w_{i+1} + \dots + w_n$ with $w_i \in W_i$ for all $i = 1, 2, \dots, n$.

Thus we get $0 + \dots + 0 + w_i + 0 + \dots + 0 = w_1 + w_2 + \dots + w_{i-1} + 0 + w_{i+1} + \dots + w_n$. Since x can be written uniquely as sum of elements x_i with $x_i \in W_i$, $w_1 = w_2 = \dots = w_{i-1} = w_i = w_{i+1} = \dots = w_n = 0$.

This implies that $x = 0$. Hence $W_i \cap \sum_{\substack{j=1 \\ j \neq i}}^n W_j = (0)$.

(ii) $\Rightarrow$ (iii). Let $w_1 + w_2 + \dots + w_n = 0$ with $w_i \in W_i$.

This implies that $w_i = -(w_1 + w_2 + \dots + w_{i-1} + w_{i+1} + \dots + w_n)$.

L.H.S. $\in W_i$ and R.H.S. $\in \sum_{\substack{j=1 \\ j \neq i}}^n W_j$.

By (ii), we get $w_i = 0$. Hence (iii) holds.

(iii) $\Rightarrow$ (v). Let $v \in V$, since $V = \sum_{i=1}^n W_i$, $v = \sum_{i=1}^n w_i \in W_i$ with $w_i \in W_i$.

Now B_i is a basis of $W_i \Rightarrow w_i = \sum_{j=1}^{k_i} \lambda_j^i b_j^i$, $\lambda_j^i \in F$

$$\Rightarrow v = \sum_{i=1}^n \sum_{j=1}^{k_i} \lambda_j^i b_j^i.$$

This shows that $\bigcup_{i=1}^n B_i$ spans V .

Now $\sum_{i=1}^n \sum_{j=1}^{k_i} \lambda_j^i b_j^i = 0$ with $\lambda_j^i \in F$

$$\Rightarrow \sum_{i=1}^n w_i = 0 \text{ where } w_i = \sum_{j=1}^{k_i} \lambda_j^i b_j^i \in W_i$$

$$\Rightarrow w_i = 0 \quad \forall \quad 1 \leq i \leq n \text{ by (i)}$$

$$\Rightarrow \lambda_j^i = 0 \quad \forall \quad j = 1, 2, \dots, k_i \text{ and } \forall i = 1, 2, \dots, n.$$

Hence $\bigcup_{i=1}^n B_i$ is a L.I. subset of V .

Consequently $\bigcup_{i=1}^n B_i$ is a basis of V .

(iv) $\Rightarrow$ (a). Let $v \in V$ can be expressed in two ways as sum of elements each one of them chosen from W_i

$$v = \sum_{i=1}^n w_i = \sum_{i=1}^n w_i' \text{ with } w_i, w_i' \in W_i$$

$$\Rightarrow \sum_{i=1}^n (w_i - w_i') = 0.$$

Now $w_i = \sum_{j=1}^{k_i} \lambda_j^i b_j^i$, $w_i' = \sum_{j=1}^{k_i} \mu_j^i b_j^i$ with $\lambda_j^i, \mu_j^i \in F$; where for all $i=1, 2, \dots, n$, $B_i = \{b_1^i, b_2^i, \dots, b_{k_i}^i\}$ is a basis of W_i .

$$\text{Then } v = \sum_{i=1}^n w_i = \sum_{i=1}^n w_i' \Rightarrow \sum_{i=1}^n \left[\sum_{j=1}^{k_i} (\lambda_j^i - \mu_j^i) b_j^i \right] = 0.$$

Since $\bigcup_{i=1}^n B_i$ is a basis of V

$$\begin{aligned} (\lambda_j^i - \mu_j^i) &= 0 \quad \forall j=1, \dots, k_i \text{ and } i=1, \dots, n. \\ \Rightarrow \lambda_j^i &= \mu_j^i \quad \forall j=1, \dots, k_i \text{ and } i=1, \dots, n \\ \Rightarrow w_i &= w_i' \quad \forall i=1, \dots, n. \end{aligned}$$

Hence $V = V_1 \oplus V_2 \oplus \dots \oplus V_n$.

Exercise 2. Let R_n be the vector space of all $n \times n$ matrices over the field of real numbers, show that if U is the set of all symmetric matrices in R_n and V is the set of all skew-symmetric matrices in R_n then U and V are subspaces of R_n such that $R_n = U \oplus V$.

Solution. Clearly O_n , the $n \times n$ zero-matrix is symmetric, so U is non-empty. Let $\alpha, \beta \in \mathbf{R}$, $X, Y \in U$ then $(\alpha X + \beta Y)^* = \alpha X^* + \beta Y^* = \alpha X + \beta Y$ as X and Y are symmetric matrices. (Note that A^* stands for transpose of matrix A .)

This means that U is a subspace of R_n .

By using the fact that a square matrix X is skew-symmetric if and only if $X^* = -X$, we get that V is also a subspace of R_n .

Now let $A \in R_n$, then $A = \frac{1}{2}(A + A^*) + \frac{1}{2}(A - A^*)$ where $[\frac{1}{2}(A + A^*)]^* = \frac{1}{2}(A^* + A)$ and $[\frac{1}{2}(A - A^*)]^* = \frac{1}{2}(A^* - A) = -\frac{1}{2}(A - A^*) \Rightarrow A = X + Y$ where $X = \frac{1}{2}(A + A^*) \in U$, $Y = \frac{1}{2}(A - A^*) \in V$, and this implies that $R_n = U + V$.

Now let $X \in U \cap V \Rightarrow X^* = X$ and $X^* = -X$

$$\Rightarrow X = -X$$

$$\Rightarrow 2X = O_n$$

$$\Rightarrow X = O_n$$

Hence

$$U \cap V = \{O_n\}$$

Thus

$$R_n = U \oplus V.$$

(As a consequence we can say that every square matrix with real entries is expressible in one and only one way as the sum of a symmetric matrix and a skew symmetric matrix.)

Exercise 3. Subspaces V_1 , V_2 and V_3 of a vector space V_F are

said to be **independent** if $V_i \cap \sum_{\substack{j=1 \\ j \neq i}}^3 V_j = (0)$ for all $i=1, 2, 3$. Show that if V_1, V_2, V_3 are **independent** subspaces then $V_i \cap V_j = (0)$ for $i \neq j, i, j=1, 2, 3$. Give an example to show that converse of this need not be true.

Solution. Since for $i \neq j, V_i \subseteq \sum_{\substack{j=1 \\ j \neq i}}^3 V_j$

$$\Rightarrow V_i \cap V_j \subseteq V_i \cap \sum_{\substack{j=1 \\ j \neq i}}^3 V_j = (0) \Rightarrow V_i \cap V_j = (0)$$

Hence if three subspaces are independent subspaces, any pair of distinct subspaces have trivial intersection.

Let $V = \mathbb{R}^3$, and let $e_1 = (1, 0, 0)$, $e_2 = (0, 1, 0)$ and $e_3 = (0, 0, 1)$. Take $V_1 = \langle e_1 \rangle$, $V_2 = \langle e_2 \rangle$, $V_3 = \langle e_1 + e_2 \rangle$.

Clearly $V_1 \cap V_2 = (0)$

Now $x \in V_1 \cap V_3 \Rightarrow x = \lambda_1 e_1 = \lambda_2 (e_1 + e_2)$ for some $\lambda_1, \lambda_2 \in \mathbb{R}$

$$\Rightarrow (\lambda_1 - \lambda_2)e_1 - \lambda_2 e_2 = 0$$

$$\Rightarrow \lambda_1 = \lambda_2, \lambda_2 = 0 \text{ as } \{e_1, e_2\} \text{ is a L.I. subset of } \mathbb{R}^3$$

$$\Rightarrow x = 0.$$

This gives that $V_1 \cap V_3 = (0)$.

Similarly it can be proved that $V_2 \cap V_3 = (0)$.

Again $V_3 \subseteq V_1 + V_2$, since $x \in V_3 \Rightarrow x = \alpha_1 (e_1 + e_2)$, $\alpha_1 \in \mathbb{R}$

$$= \alpha_1 e_1 + \alpha_1 e_2 \Rightarrow x \in V_1 + V_2.$$

Consequently $V_3 \cap (V_1 + V_2) = V_3 \neq (0)$.

Thus the subspaces V_1, V_2 and V_3 are not independent.

PROBLEMS

1. Consider the vector space $V = \mathbb{R}^3$. Determine complements of the subspaces of V generated by the following subsets of V

(a) $\{(1, 0, 0), (0, 1, 0)\}$

(b) $\{(1, -1, 3), (-4, 2, 0), (-3, -1, -15)\}$

(c) $\{(1, 1, 0), (0, 1, 0)\}$

(d) For fixed and distinct real numbers α and β ,

$$\{(1, \alpha, \alpha^2), (1, \beta, \beta^2)\}.$$

[Ans. Note that complement of a given space is not unique. In the following we give one complement of each of the subspaces generated by the subset (a), (b), (c) and (d). (a), $\langle (0, 0, 1) \rangle$ (b), $\{0\}$; (c), $\langle (0, 0, 1) \rangle$; (d), $\langle (0, 0, 1) \rangle$.]

2. Let V_F be a vector space with a basis B . If B_1 is any subset

of B , and W is the subspace generated by B_1 , show that the subspace of V generated by $B - B_1$, is a complement of W .

3. Show that the vectors $(1, 0, 1)$, $(2, 1, 4)$ are L.I. in $\mathbf{R}^3$. Find a basis of $\mathbf{R}^3$ containing these vectors.

Use the same basis to find a complement of the subspace $\langle (1, 0, 1), (2, 1, 4) \rangle$ in $\mathbf{R}^3$.

[Ans. $\{(1, 0, 1), (2, 1, 4), (0, 0, 1)\}, \langle (0, 0, 1) \rangle$].

4. Let F be a field having q elements and let V_F be a vector space of dimension n . Let W be a subspace of V_F of dimension $m < n$. Find how many complements W has in V ?

[Ans. $q^{m(n-m)}$]
5. Let V_F be a vector space of dimension n . If V_1 and V_2 are subspaces each of dimension strictly greater than $n/2$ then prove that $V_1 \cap V_2 \neq (0)$.

6. Let x, y, z be elements of a vector space V_F . Let L, M and N be the subspaces spanned by x, y and z respectively show that x, y and z are L.I. over F if and only if L, M and N are independent.

7. Let V be the vector space of all real valued functions on $\mathbf{R}$. If $V_1 = \{f \in V \mid f(-x) = f(x)\}$ and $V_2 = \{f \in V \mid f(-x) = -f(x)\}$. Show that V_1 and V_2 are subspaces of V and $V = V_1 \oplus V_2$.

8. V_1 and V_2 are subspaces of a vector space V such that $V = V_1 + V_2$. If W is a complement of $V_1 \cap V_2$ in V_2 [i.e. $V_2 = (V_1 \cap V_2) \oplus W$], prove that $V = V_1 \oplus W$.

9. V is a finite-dimensional vector space over a field F . Let $V = V_1 \oplus V_2$ and let L be any subspace of V . Prove that $2 \dim L - \dim V \leq \dim [V_1 \cap L] \oplus (V_2 \cap L) \leq \dim L$.

10. Show by means of an example that under certain circumstances, it is possible to find three subspaces V_1, V_2, V_3 of a finite-dimensional vector space V_F such that $V = V_1 \oplus V_2 = V_2 \oplus V_3 = V_3 \oplus V_1$. What does this imply about the dimension of V ?

[Ans. Dimension of V is an even integer]

6. Matrices and Change of Basis

Definition 11.34. (Ordered Basis). Let V_F be any vector space of finite dimension n .

Then any ordered n -tuple $(x_1, x_2, \dots, x_n)$ of n members of V_F is called an *ordered basis* of V_F if the set $\{x_1, x_2, \dots, x_n\}$ is a basis of V_F .

Thus in an ordered basis, the order of arrangement of members of the basis is also taken into account. For example is $V = F^{(3)}$, we know that $\{e_1, e_2, e_3\}$ where $e_1 = (1, 0, 0)$, $e_2 = (0, 1, 0)$, $e_3 = (0, 0, 1)$ is a basis of V . Then each of ordered triples (e_1, e_2, e_3) , (e_2, e_1, e_3) and (e_1, e_3, e_2) is an ordered basis of V_F . Note that as these three ordered

triples are distinct as ordered bases, so these are three different ordered bases of V_F .

This concept of ordered basis is very useful in the discussion of matrices. This we intend to do in this section.

Let $(x_1, x_2, \dots, x_n)$ be a *fixed* ordered basis of a vector space V_F . Let $(y_1, y_2, \dots, y_r)$ be any r -tuple of vectors in V_F . As each element of V is expressible uniquely as a linear combination of $x_1, x_2, \dots, x_n$ we have

$$y_j = \sum_{i=1}^n \alpha_{ij} x_i; j=1, 2, \dots, r$$

with α_{ij} (uniquely determined) elements of F . Thus we get an $n \times r$ matrix over F .

$$(\alpha_{ij}) = \begin{bmatrix} \alpha_{11} & \alpha_{12} & \dots & \dots & \alpha_{1r} \\ \alpha_{21} & \alpha_{22} & \dots & \dots & \alpha_{2r} \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ \alpha_{n1} & \alpha_{n2} & \dots & \dots & \alpha_{nr} \end{bmatrix}$$

We call this matrix, *the matrix of $(y_1, y_2, \dots, y_r)$ relative to or with respect to $(x_1, x_2, \dots, x_n)$* .

Recall that an $n \times n$ matrix (α_{ij}) over a field F is said to be **non-singular** if there exists an $n \times n$ matrix (β_{ij}) over F such that $(\alpha_{ij})(\beta_{ij}) = I = (\beta_{ij})(\alpha_{ij})$, where I is the identity matrix over F .

Now we prove the following:

Theorem 11.35. *Let $(x_1, x_2, \dots, x_n)$ be an ordered basis of V_F and $(y_1, y_2, \dots, y_n)$ be any ordered n -tuple of elements of V_F . Then $(y_1, y_2, \dots, y_n)$ is ordered basis of V_F if and only if the matrix of $(y_1, y_2, \dots, y_n)$ relative to $(x_1, x_2, \dots, x_n)$ is non-singular.*

Proof. Let (α_{ij}) be the matrix of $(y_1, y_2, \dots, y_n)$ relative to $(x_1, x_2, \dots, x_n)$. By definition

$$y_j = \sum_{i=1}^n \alpha_{ij} x_i; j=1, 2, \dots, n \quad \dots(1)$$

Let $(y_1, y_2, \dots, y_n)$ be an ordered basis of V_F and let (β_{ij}) be the matrix of $(x_1, x_2, \dots, x_n)$ relative to $(y_1, y_2, \dots, y_n)$. Then

$$x_i = \sum_{k=1}^n \beta_{ki} y_k; i=1, 2, \dots, n. \quad \dots(2)$$

From (1) and (2) we get

$$\begin{aligned} x_i &= \sum_{k=1}^n \beta_{ki} \left(\sum_{l=1}^n \alpha_{lk} x_l \right) \\ &= \sum_{l=1}^n \left(\sum_{k=1}^n \alpha_{lk} \beta_{ki} \right) x_l. \end{aligned} \quad \dots(3)$$

However $x_i = \sum_{l=1}^n \delta_{li} x_l$ where $\delta_{li}=0$ for $l \neq i$ and $\delta_{li}=1$ for $l=i$.

$$\text{Thus } \sum_{l=1}^n \delta_{li} x_l = \sum_{l=1}^n \left(\sum_{k=1}^n \alpha_{lk} \beta_{ki} \right) x_l$$

$$\text{This gives that } \delta_{li} = \sum_{k=1}^n \alpha_{lk} \beta_{ki}$$

$$\text{So that } (\delta_{li}) = (\alpha_{ij}) (\beta_{ij})$$

$$\text{i.e. } I = (\alpha_{ij}) (\beta_{ij}).$$

$$\text{Similarly } (\beta_{ij})(\alpha_{ij}) = I.$$

Hence (α_{ij}) is a non-singular matrix.

Conversely let (α_{ij}) be non-singular and (β_{ij}) be $(\alpha_{ij})^{-1}$. Then it follows that for any $i=1, 2, \dots, n$

$$\begin{aligned} \sum_{j=1}^n \beta_{ji} y_j &= \sum_{j=1}^n \beta_{ji} \left(\sum_{k=1}^n \alpha_{kj} x_k \right) \\ &= \sum_{k=1}^n \left(\sum_{j=1}^n \alpha_{kj} \beta_{ji} \right) x_k \\ &= x_i, \text{ since } \sum_{j=1}^n \alpha_{kj} \beta_{ji} = 0 \text{ if } k \neq i \\ &= 1 \text{ if } k = i. \end{aligned}$$

Hence each $x_i \in \langle y_1, y_2, y_3, \dots, y_n \rangle$.

So that $V \langle y_1, y_2, \dots, y_n \rangle$. This implies $y_1, y_2, \dots, y_n$ is a basis of V , since $\dim V = n$.

Hence $(y_1, y_2, \dots, y_n)$ is an ordered basis of V .

This completes the theorem. ■

Definition 11.36. The ordered basis $(e_1, e_2, \dots, e_n)$ of $F^{(n)}$ where $e_1 = (1, 0, 0, \dots, 0)$, $e_2 = (0, 1, 0, \dots, 0)$, ..., $e_n = (0, 0, \dots, 1)$ is called the *standard basis* of $F^{(n)}$.

EXAMPLE 15. Consider $F^{(3)}$. Consider $f_1 = (1, 0, 0)$, $f_2 = (1, \alpha, 0)$, $f_3 = (1, \alpha, \beta)$ where α, β are non-zero members of F .

It can be easily seen that these vectors are L.I. So that (f_1, f_2, f_3) is an ordered basis of V .

Let (e_1, e_2, e_3) be the standard basis of V .

Since $e_1 = (1, 0, 0)$, $e_2 = (0, 1, 0)$, $e_3 = (0, 0, 1)$ we get

$$f_1 = e_1 = 1.e_1 + 0.e_2 + 0.e_3$$

$$f_2 = 1.e_1 + \alpha.e_2 + 0.e_3$$

$$\text{and } f_3 = 1.e_1 + \alpha.e_2 + \beta.e_3.$$

Thus the matrix of (f_1, f_2, f_3) relative to (e_1, e_2, e_3) is

$$A = \begin{pmatrix} 1 & 1 & 1 \\ 0 & \alpha & \alpha \\ 0 & 0 & \beta \end{pmatrix}$$

Since

$$\begin{aligned} e_1 &= f_1 = 1 \cdot f_1 + 0 \cdot f_2 + 0 \cdot f_3 \\ e_2 &= -\alpha^{-1} \cdot f_1 + \alpha^{-1} f_2 + 0 \cdot f_3 \\ e_3 &= 0 \cdot f_1 - \beta^{-1} \cdot f_2 + \beta^{-1} \cdot f_3 \end{aligned}$$

we get that the matrix of (e_1, e_2, e_3) relative to (f_1, f_2, f_3) is

$$B = \begin{pmatrix} 1 & -\alpha^{-1} & 0 \\ 0 & \alpha^{-1} & -\beta^{-1} \\ 0 & 0 & \beta^{-1} \end{pmatrix}$$

Then $B = A^{-1}$.

Theorem 11.37. Let V_F be a vector space of dimension n . If $A = (\alpha_{ij})$ is a matrix of the basis $(f_1, f_2, \dots, f_n)$ with respect to basis $(e_1, e_2, \dots, e_n)$ and $B = (\beta_{ij})$ is the matrix of the basis $(g_1, g_2, \dots, g_n)$ with respect to $(f_1, f_2, \dots, f_n)$, then the matrix of $(g_1, g_2, \dots, g_n)$ with respect to $(e_1, e_2, \dots, e_n)$ is AB .

Proof: Now $f_k = \sum_{j=1}^n \alpha_{jk} e_j$; $k=1, 2, \dots, n$

$$g_i = \sum_{k=1}^n \beta_{ki} f_k; i=1, 2, \dots, n.$$

Thus $g_i = \sum_{k=1}^n \beta_{ki} \left(\sum_{j=1}^n \alpha_{jk} e_j \right).$

$$= \sum_{j=1}^n \left(\sum_{k=1}^n \alpha_{jk} \beta_{ki} \right) e_j$$

$$= \sum_{j=1}^n \gamma_{ji} e_j; i=1, 2, \dots, n$$

where $\gamma_{ji} = \sum_{k=1}^n \alpha_{jk} \beta_{ki}$, So that $(\gamma_{ij}) = (\alpha_{ij})(\beta_{ij})$.

Hence the matrix of $(g_1, g_2, \dots, g_n)$ relative to $(e_1, e_2, \dots, e_n)$ is

$$(\gamma_{ij}) = (\alpha_{ij})(\beta_{ij}) = AB.$$

This proves the theorem. ■

PROBLEMS

1. Let V_F be a vector space of dimension n and let $(e_1, e_2, \dots, e_n)$ be an ordered basis of V_F . Prove that

$$(f_1, f_2, \dots, f_n) \leftrightarrow (\alpha_{ij})$$

where $(f_1, f_2, \dots, f_n)$ is an ordered basis of V_F and (α_{ij}) is the matrix of $(f_1, f_2, \dots, f_n)$ relative to $(e_1, e_2, \dots, e_n)$, is a 1-1 correspondence between the set of all ordered bases of V_F and the set of all $n \times n$ non-singular matrices over F .

2. Let $V = \mathbb{R}^3$. Find the matrices of the ordered basis (f_1, f_2, f_3) relative to the standard basis (e_1, e_2, e_3) of V_F when f_1, f_2, f_3 are as given below:

$$(i) f_1 = (1, \cos x, \sin x), f_2 = (1, 0, 0), f_3 = (1, -\sin x, \cos x)$$

$$(ii) f_1 = (2, 1, 0), f_2 = (0, 2, 1), f_3 = (0, 1, 2).$$

In each case find the matrix of (e_1, e_2, e_3) relative to (f_1, f_2, f_3) .

[Ans. Matrices of (f_1, f_2, f_3) relative (e_1, e_2, e_3) are

$$(i) \begin{pmatrix} 1 & 1 & 1 \\ \cos x & 0 & -\sin x \\ \sin x & 0 & \cos x \end{pmatrix} \quad (ii) \begin{pmatrix} 2 & 0 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 2 \end{pmatrix}$$

Matrices of (e_1, e_2, e_3) relative to (f_1, f_2, f_3) are

$$(i) \begin{pmatrix} 0 & \cos x & \sin x \\ 1 & \sin x - \cos x & -\sin x - \cos x \\ 0 & -\sin x & \cos x \end{pmatrix}$$

$$(ii) \left[\begin{pmatrix} \frac{1}{2} & 0 & 0 \\ -\frac{1}{3} & \frac{2}{3} & -\frac{1}{3} \\ \frac{1}{6} & -\frac{1}{3} & \frac{2}{3} \end{pmatrix} \right]$$

3. Find the total number of ordered bases of $V = F^{(n)}$ where F is a field containing q elements. [Ans. $(q^n - 1)(q^n - q)(q^n - q^{n-1})$]

7. Inner Product Spaces

The properties of vector spaces discussed in earlier sections mainly depended on addition of vectors and multiplication of vectors by scalars. In this section we introduce another structure in a vector space by postulating the existence of functions like inner product and norm. These functions generalize the concepts of length of a vector, perpendicularity between two vectors and angle between two vectors. At the outset this adventure looks a misfit and of no immediate use, but in fact it provides a link between analysis and algebra. A vector space equipped with a 'norm' becomes a metric space and so one can utilize the concepts of limits, sequences and continuity. However, we shall not be able to explore all these possibilities, instead we shall

concentrate on the algebraic aspects of these functions. To develop the theory in a reasonable manner (as to make it look a natural extension of geometrical notions of length, perpendicularity and angle) we have to sacrifice some generality. So in this section, we no longer consider vector spaces V over an arbitrary field F ; rather we restrict F to be the field of *real* or *complex* numbers. In the former case V_F is called a *real vector space* and in the latter case, a *complex vector space*.

REMARK. If $\alpha \in \mathbf{R}$ then $|\alpha|$ stands for absolute value of α i.e. $|\alpha| = \alpha$ for $\alpha \geq 0$ and $|\alpha| = -\alpha$ for $\alpha < 0$. Further if $\alpha = x + iy$, $x, y \in \mathbf{R}$ then $|\alpha|$ stands for *modulus* of α i.e. $|\alpha| = +\sqrt{x^2 + y^2}$.

Definition 11.38. A vector space V_F is called an *inner product space* if there is a function $f: V \times V \rightarrow F$ satisfying the following axioms.

For $u, v, w \in V$; $\alpha, \beta \in F$.

IP-1. $f(u, v) = \overline{f(v, u)}$ i.e. the complex conjugate of $f(v, u)$.

IP-2. $f(u, u) \geq 0$ and $f(u, u) = 0 \Leftrightarrow u = 0$.

IP-3. $f(\alpha u + \beta v, w) = \alpha f(u, w) + \beta f(v, w)$.

Function f satisfying IP-1, IP-2, IP-3 is called an *inner product* on V .

REMARK. 1. Because of IP-1, $f(u, u) = \overline{f(u, u)}$. So $f(u, u)$ is real and hence IP-2 makes sense.

2. If $F = \mathbf{R}$ then IP-1 merely states that $f(u, v) = f(v, u)$.

NOTATION. $f(u, v)$ is generally denoted by any one of the following symbols

$$(u, v); u \cdot v; \langle u | v \rangle.$$

In the sequel we shall use the third symbol.

EXAMPLE 16. Let $u = (a, b)$, $v = (c, d) \in \mathbf{C}^2$

Define $\langle u | v \rangle$ as $a\bar{c} + b\bar{d}$. We claim that it is an inner-product on $\mathbf{C}^2$. We verify all the three axioms.

IP-1 $\langle v | u \rangle = \overline{\langle u | v \rangle} = \overline{a\bar{c} + b\bar{d}} = \bar{a}c + \bar{b}d = \langle u | v \rangle$.

IP-2 $\langle u | u \rangle = a\bar{a} + b\bar{b} = |a|^2 + |b|^2 \geq 0$ and $\langle u | u \rangle = 0 \Leftrightarrow |a|^2 + |b|^2 = 0 \Leftrightarrow |a| = |b| = 0 \Leftrightarrow a = b = 0 \Leftrightarrow u = 0$.

IP-3. Let $\alpha, \beta \in \mathbf{C}$ and $w = (x, y) \in \mathbf{C}^2$.

Then $\langle \alpha u + \beta v | w \rangle = (\alpha a + \beta c)\bar{x} + (\alpha b + \beta d)\bar{y}$
 $= \alpha(a\bar{x} + b\bar{y}) + \beta(c\bar{x} + d\bar{y}) = \alpha \langle u | w \rangle + \beta \langle v | w \rangle$.

Hence $\mathbf{C}^2$ is an inner product space.

Note. The inner product defined in example 16 is usually referred to as the *standard inner product*.

EXAMPLE 17. Let $u = (a, b)$, $v = (c, d) \in \mathbf{R}^2$, define

$$\langle u | v \rangle = ac - bc - ad + 4bd.$$

Clearly $\langle v | u \rangle = ca - cb - da + 4db = \langle u | v \rangle$.

Also $\langle u | u \rangle = a^2 - ab - ab + 4b^2 = a^2 - 2ab + 4b^2 = (a-b)^2 + 3b^2 \geq 0$ and $\langle u | u \rangle = 0 \Leftrightarrow (a-b)^2 + 3b^2 = 0 \Leftrightarrow a=b=0 \Leftrightarrow u=0$.

Finally if $w = (e, f) \in \mathbf{R}^2$ and $\alpha, \beta \in \mathbf{R}$, then

$$\begin{aligned} \langle \alpha u + \beta v | w \rangle &= (\alpha a + \beta c)e - (\alpha b + \beta d)e - (\alpha a + \beta c)f + 4(\alpha b + \beta d)f \\ &= \alpha(ae - be - af + 4bf) + \beta(ce - de - cf + 4df) \\ &= \alpha \langle u | w \rangle + \beta \langle v | w \rangle. \end{aligned}$$

Hence $\mathbf{R}^2$ is an inner product space.

EXAMPLE 18. Let V be the vector space over $\mathbf{C}$ of all continuous complex-valued functions on the interval $[0, 1]$. For $f(t), g(t) \in V$, define $\langle f(t) | g(t) \rangle$ by $\int_0^1 f(t) \overline{g(t)} dt$.

$$\begin{aligned} \text{Evidently } \int_0^1 g(t) \overline{f(t)} dt &= \int_0^1 \overline{g(t)} f(t) dt \\ &= \int_0^1 f(t) \overline{g(t)} dt. \end{aligned}$$

Thus we get $\langle f(t) | g(t) \rangle = \overline{\langle g(t) | f(t) \rangle}$.

Again $\langle f(t) | f(t) \rangle = \int_0^1 f(t) \overline{f(t)} dt = \int_0^1 |f(t)|^2 dt \geq 0$ and the equality holds if and only if $f(t) = 0$.

Finally for $h(t) \in V$ and $\alpha, \beta \in \mathbf{C}$,

$$\begin{aligned} \langle \alpha f(t) + \beta g(t) | h(t) \rangle &= \int_0^1 [\alpha f(t) + \beta g(t)] \overline{h(t)} dt \\ &= \alpha \int_0^1 f(t) \overline{h(t)} dt + \beta \int_0^1 g(t) \overline{h(t)} dt \\ &= \alpha \langle f(t) | h(t) \rangle + \beta \langle g(t) | h(t) \rangle. \end{aligned}$$

Hence V is an inner product space.

EXAMPLE 29. Consider the vector space $\mathbf{R}^n$.

For $u = (x_1, x_2, \dots, x_n), v = (y_1, y_2, \dots, y_n) \in \mathbf{R}^n$, put

$$\langle u | v \rangle = \sum_{i=1}^n x_i y_i.$$

It can be verified that all the axioms IP-1, IP-2 and IP-3 are satisfied and so $\mathbf{R}^n$ is an inner product space.

Note: Henceforth, throughout the remainder of this section V will denote an inner product space.

Definition 11.39. A finite-dimensional real inner product space is called a *Euclidean Space* while a finite-dimensional complex inner product space is called a *Unitary Space*. Before we proceed to the next definition, we shall like to impress upon the readers that henceforth whenever we talk of square root of a real non-negative number,

we shall always mean the *positive* value of the square root. Thus, according to this convention $\sqrt{4}=2$ not -2 .

Definition 11.40. For any $v \in V$, the *norm* of v (or the *length* of v) is defined to be $\sqrt{\langle v | v \rangle}$.

Remark : Thus norm of $v \in V$ is always a non-negative real number.

NOTATION : Norm of v will be denoted by $\|v\|$.

Definition 11.41. For $u, v \in V$, $\|u - v\|$ is called the *distance* between u and v .

NOTATION : Distance between u and v will be denoted by $d(u, v)$.

We pause here to prove certain properties of inner product spaces, which are direct consequences IP-1, of IP-2, IP-3 and definition 11.40.

Theorem 11.42. In an inner-product space V ; for all $u, v, w, x \in V$, $\alpha, \beta, \gamma, \delta \in F$

$$(i) \langle u | \alpha v + \beta w \rangle = \bar{\alpha} \langle u | v \rangle + \bar{\beta} \langle u | w \rangle.$$

$$(ii) \langle \alpha u + \beta v | \gamma w + \delta x \rangle = \alpha \bar{\gamma} \langle u | w \rangle + \alpha \bar{\delta} \langle u | x \rangle + \beta \bar{\gamma} \langle v | w \rangle + \beta \bar{\delta} \langle v | x \rangle.$$

$$(iii) \langle 0 | v \rangle = \langle u | 0 \rangle = 0.$$

$$(iv) \langle u | v \rangle = 0 \forall v \in V \Rightarrow u = 0 \text{ and}$$

$$\langle u | v \rangle = 0 \forall u \in V \Rightarrow v = 0$$

$$(v) \|\alpha u\| = |\alpha| \|u\|.$$

Proof : (i) $\langle u | \alpha v + \beta w \rangle = \langle \alpha v + \beta w | u \rangle$ by IP-1.

$$= \alpha \langle v | u \rangle + \beta \langle w | u \rangle \text{ by IP-3}$$

$$= \bar{\alpha} \langle v | u \rangle + \bar{\beta} \langle w | u \rangle, \text{ property of complex conjugation.}$$

$$= \bar{\alpha} \langle u | v \rangle + \bar{\beta} \langle u | w \rangle \text{ by IP-1}$$

$$(ii) \langle \alpha u + \beta v | \gamma w + \delta x \rangle = \alpha \langle u | \gamma w + \delta x \rangle + \beta \langle v | \gamma w + \delta x \rangle$$

by IP-3

$$= \alpha \bar{\gamma} \langle u | w \rangle + \alpha \bar{\delta} \langle u | x \rangle + \beta \bar{\gamma} \langle v | w \rangle + \beta \bar{\delta} \langle v | x \rangle$$

by (i)

$$= \alpha \bar{\gamma} \langle u | w \rangle + \alpha \bar{\delta} \langle u | x \rangle + \bar{\beta} \gamma \langle v | w \rangle + \bar{\beta} \delta \langle v | x \rangle.$$

$$(ii) \langle 0 | v \rangle = \langle 0u + 0w | v \rangle = 0 \langle u | v \rangle + 0 \langle w | v \rangle = 0.$$

$$\text{Similarly } \langle u | 0 \rangle = \langle u | 0v + 0w \rangle = 0 \langle u | v \rangle + 0 \langle u | w \rangle = 0.$$

$$(iv) \langle u | v \rangle = 0 \forall v \in V \Rightarrow \langle u | u \rangle = 0 \Rightarrow u = 0 \text{ by IP-2.}$$

$$\text{Similarly } \langle u | v \rangle = 0 \forall u \in V \Rightarrow \langle v | v \rangle = 0 \Rightarrow v = 0 \text{ by IP-2.}$$

$$(v) \|\alpha u\|^2 = \langle \alpha u | \alpha u \rangle = \alpha \bar{\alpha} \langle u | u \rangle \text{ by (ii)}$$

$$= |\alpha|^2 \|u\|^2.$$

Hence

$$\|\alpha u\| = |\alpha| \|u\|. \blacksquare$$

Theorem 11.43. (Schwarz Inequality).

For all $u, v \in V$, $|\langle u | v \rangle| \leq \|u\| \|v\|$.

Proof: If $u=0$ then left hand side $= \langle 0 | v \rangle = 0$ and right hand side $= \|0\| \|v\| = 0$, so $\langle u | v \rangle = \|u\| \|v\|$.

Assume now that $u \neq 0$ then $\|u\| \neq 0$, else we shall get $\|u\| = 0 \Rightarrow \langle u | u \rangle = 0 \Rightarrow u = 0$, a contradiction to the assumption.

$$\text{Put } w = v - \frac{\langle v | u \rangle}{\|u\|^2} u.$$

Since $\|w\|^2 \geq 0$, we get $\langle v - \frac{\langle v | u \rangle}{\|u\|^2} u | v - \frac{\langle v | u \rangle}{\|u\|^2} u \rangle \geq 0$

$$\Rightarrow \langle v | v \rangle - \frac{\langle v | u \rangle}{\|u\|^2} \langle v | u \rangle - \frac{\langle v | u \rangle}{\|u\|^2} \langle u | v \rangle + \frac{\langle v | u \rangle \langle v | u \rangle}{\|u\|^2} \langle u | u \rangle \geq 0$$

$$\Rightarrow \|v\|^2 - \frac{|\langle v | u \rangle|^2}{\|u\|^2} - \frac{\langle u | v \rangle \langle u | v \rangle}{\|u\|^2} + \frac{|\langle v | u \rangle|^2}{\|u\|^4} \|u\|^2 \geq 0$$

$$\Rightarrow \|v\|^2 - \frac{|\langle u | v \rangle|^2}{\|u\|^2} \geq 0$$

$$\Rightarrow |\langle u | v \rangle|^2 \leq \|u\|^2 \|v\|^2$$

$$\Rightarrow |\langle u | v \rangle| \leq \|u\| \|v\|. \blacksquare$$

REMARK. This inequality is also called Cauchy-Schwarz inequality.

Theorem 11.44. Every inner product space is a metric space.

Proof: Let V be an inner-product space.

For any $u, v \in V$, put $d(u, v) = \|u - v\|$.

Since norm is always non-negative we get $d(u, v) \geq 0$.

Further $d(u, v) = 0 \Leftrightarrow \|u - v\| = 0 \Leftrightarrow \langle (u - v) | (u - v) \rangle = 0 \Leftrightarrow u = v$.

Also $d(u, v) = \|u - v\| = \sqrt{\langle (u - v) | (u - v) \rangle}$

$$= \sqrt{(-1)(-1) \langle (v - u) | (v - u) \rangle}$$

$$= \sqrt{\langle (v - u) | (v - u) \rangle} = \|v - u\| = d(v, u).$$

We claim that $\|u + v\| \leq \|u\| + \|v\|$ for all $u, v \in V$.

Now $\|u + v\|^2 = \langle u + v | u + v \rangle$

$$= \langle u | u \rangle + \langle v | u \rangle + \langle u | v \rangle + \langle v | v \rangle$$

$$= \|u\|^2 + \langle u | v \rangle + \langle u | v \rangle + \|v\|^2$$

$$= \|u\|^2 + 2R\langle u | v \rangle + \|v\|^2, \text{ where } R(\alpha) \text{ stands for}$$

real part of complex number α ,

As $R(\alpha) \leq |\alpha|$, we get

$$\|u + v\|^2 \leq \|u\|^2 + 2|\langle u | v \rangle| + \|v\|^2$$

$\leq \|u\|^2 + 2\|u\| \|v\| + \|v\|^2$ by Schwarz inequality

$$\Rightarrow \|u+v\|^2 \leq (\|u\| + \|v\|)^2$$

$\Rightarrow \|u+v\| \leq \|u\| + \|v\|$ and our claim is established.

Finally for $u, v, w \in V$

$$d(u, v) + d(v, w) = \|u-v\| + \|v-w\|$$

$$\geq \|u-v+v-w\| \text{ by our claim}$$

$$\Rightarrow d(u, v) + d(v, w) \geq \|u-w\|, \text{ i.e. } d(u, w).$$

Hence V is a metric space. ■

In geometry the concept of perpendicularity is of much importance. In the following we introduce a similar concept in inner product spaces.

Definition 11.46 $u \in V$ is said to be *orthogonal* to $v \in V$ if $\langle u | v \rangle = 0$.

REMARK. By definition it follows that u is orthogonal to v implies that v is orthogonal to u . Since

$$\langle u | v \rangle = 0 \Rightarrow \overline{\langle v | u \rangle} = 0 \Rightarrow \langle v | u \rangle = \bar{0} = 0.$$

Definition 11.45. If W is a subspace of V , then an orthogonal complement of W (denoted by $W^\perp$) is defined by

$$\{u \in V \mid \langle u | w \rangle = 0 \forall w \in W\}.$$

EXAMPLE 20. As seen in Example 16, if $u = (a, b)$, $v = (c, d) \in \mathbb{C}^2$, then $\langle u | v \rangle = a\bar{c} + b\bar{d}$ is an inner product. The vectors $x = (3, 5)$ and $y = (5i, -3i)$ are orthogonal. Since $\langle x | y \rangle = 3(-5i) + 5(3i) = 0$

EXAMPLE 21. In Example 18, the vectors $\sin \pi t$ and $\cos \pi t$ are orthogonal, since $\langle \sin \pi t | \cos \pi t \rangle = \int_0^1 \sin \pi t \cos \pi t dt$

$$= \frac{1}{2} \int_0^1 \sin 2\pi t dt = -\frac{1}{4\pi} (\cos 2\pi t)_0^1 = 0.$$

EXAMPLE 22. Let $V = \mathbb{R}^2$ and $W = \{(a, a) \mid a \in \mathbb{R}\}$. Then W is a subspace of V . We equip V by an inner-product as defined in Example 17. That is for all $u = (x, y)$, $v = (x', y') \in \mathbb{R}^2$, $\langle u | v \rangle = xx' - yy' - xy' + 4x'y'$.

Then we claim that $W^\perp = \{(x, 0) \mid x \in \mathbb{R}\}$.

Since $(p, q) \in W^\perp \Rightarrow \langle (a, a) | (p, q) \rangle = 0$ for all $a \in \mathbb{R}$.

Let $a \neq 0$ then $(p, q) \in W^\perp \Rightarrow ap - ap - aq + 4aq = 0 \Rightarrow 3aq = 0 \Rightarrow q = 0$.

Also for any $r \in \mathbb{R}$, $\langle (a, a) | (r, 0) \rangle = ar - ar - 0 + 4 \cdot 0 = 0 \forall (a, a) \in W$.

Hence our claim is established.

Lemma 11.47. For any subspace W of V , $W^\perp$ is a subspace of V such

that $W \cap W^\perp = (0)$.

Proof : Obviously $0 \in W^\perp$ so $W^\perp$ is non-empty.

Let $u, v \in W^\perp$ and $\alpha, \beta \in F$. Then $\langle u | w \rangle = \langle v | w \rangle = 0$ for all $w \in W^\perp$. Now for all $w \in W$, $\langle \alpha u + \beta v | w \rangle = \alpha \langle u | w \rangle + \beta \langle v | w \rangle = 0 \Rightarrow \alpha u + \beta v \in W^\perp$.

Hence $W^\perp$ is a subspace of V .

Finally $x \in W \cap W^\perp \Rightarrow x \in W$ and $x \in W^\perp \Rightarrow \langle x | x \rangle = 0 \Rightarrow x = 0 \Rightarrow W \cap W^\perp = (0)$. ■

Definition 11.48. Let $v_i \in V$ for all $i \in \Lambda$, Λ being any index set. $\{v_i\}_{i \in \Lambda}$ is called an *orthogonal set* if for all $i, j \in \Lambda, i \neq j$, $\langle v_i | v_j \rangle = 0$.

Definition 11.49. Let $v_i \in V$ for all $i \in \Lambda$, Λ being any index set. Then $\{v_i\}_{i \in \Lambda}$ is called an *orthonormal set* if

(i) for each $i \in \Lambda$, $\|v_i\| = 1$,

(ii) for all $i, j \in \Lambda, i \neq j$, $\langle v_i | v_j \rangle = 0$.

In this way an orthonormal set is an orthogonal set in which every vector is of norm 1.

EXAMPLE 23. Let V be the vector space over $\mathbf{R}$ of all real valued continuous functions defined on $[0, 1]$ with the inner product

$$\langle f | g \rangle = \int_0^1 f(x)g(x) dx \text{ for all } f, g \in V.$$

For all $n \in \mathbf{N}$, define $f_n(x) = \sqrt{2} \cos 2\pi nx$

and $g_n(x) = \sqrt{2} \sin 2\pi nx$.

Consider $X = \{1, f_1, g_1, f_2, g_2, \dots\}$.

Clearly $\forall n \in \mathbf{N}$, $\langle f_n | f_n \rangle = \int_0^1 2 \cos^2 2\pi nx dx$

$$\begin{aligned} & \int_0^1 (1 + \cos 4\pi nx) dx \\ &= 1 + \frac{1}{4\pi n} \left(\sin 4\pi nx \right) \Big|_0^1 = 1 \Rightarrow \|f\| = 1, \end{aligned}$$

Similarly $\forall n \in \mathbf{N}$, $\langle g_n | g_n \rangle = 1 \Rightarrow \|g_n\| = 1$.

Further for $n \neq m$, (i) $\langle f_n | g_m \rangle = \int_0^1 2 \cos 2\pi nx \sin 2\pi mx dx$

$$= \int_0^1 [\sin 2\pi(n+m)x - \sin 2\pi(n-m)x] dx = 0$$

(ii) $\langle f_n | f_m \rangle = \int_0^1 2 \cos 2\pi nx \cos 2\pi mx dx$

$$\begin{aligned} &= \int_0^1 [\cos 2\pi(n+m)x + \cos 2\pi(n-m)x] dx \\ &= 0. \end{aligned}$$

$$\begin{aligned}
 \text{(iii)} \quad \langle g_n | g_m \rangle &= \int_0^1 2 \sin 2\pi n x \sin 2\pi m x \, dx \\
 &= \int_0^1 [\cos 2\pi(m-n)x - \cos 2\pi(n+m)x] \, dx \\
 &= 0
 \end{aligned}$$

$$\text{Also for all } n \in N \quad \langle 1 | f_n \rangle = \int_0^1 \sqrt{2} \cos 2\pi n x \, dx = 0$$

$$\text{and} \quad \langle 1 | g_n \rangle = \int_0^1 \sqrt{2} \sin 2\pi n x \, dx = 0.$$

Hence X is an orthonormal subset of V .

EXAMPLE 24. Consider $\mathbb{C}^2$ with the standard inner product. Then $\{e_1, e_2\}$ is an orthonormal subset of $\mathbb{C}^2$ where $e_1 = (1, 0)$ and $e_2 = (0, 1)$, since $\langle e_1 | e_1 \rangle = 1\bar{1} + 0\bar{0} = 1$ and $\langle e_2 | e_2 \rangle = 0\bar{0} + 1\bar{1} = 1 \Rightarrow \|e_1\| = \|e_2\| = 1$ and $\langle e_1 | e_2 \rangle = 1\bar{0} + 0\bar{1} = 0$.

Lemma 11.50 Any orthogonal set of non-zero vectors is linearly independent.

Proof: Let X be an orthogonal subset of an inner product space V with the property that $0 \notin X$. Let $v_1, v_2, \dots, v_k \in X$ be such that

$$\sum_{i=1}^k \alpha_i v_i = 0, \quad \alpha_i \in F. \text{ For any } j \in \{1, 2, 3, \dots, k\}, \quad \langle \sum_{i=1}^k \alpha_i v_i | v_j \rangle = 0.$$

$$\Rightarrow \alpha_j \langle v_j | v_j \rangle = 0. \text{ Since } \langle v_i | v_j \rangle = 0 \text{ for } i \neq j.$$

$$\Rightarrow \alpha_j = 0 \text{ as } \langle v_j | v_j \rangle \neq 0 \text{ (recall that } v_j \neq 0).$$

Hence $v_1, v_2, \dots, v_k$ are L.I. over F .

As a consequence X is a L. I. subset of V . ■

Corollary 11.51. If $\{v_j\}$ is an orthonormal subset of V , then the vectors v_i 's are L.I. over F . Further if $w = \sum_{i=1}^k \alpha_i v_i$ then for each

$$i \in \{1, 2, 3, \dots, k\}, \quad \alpha_i = \langle w | v_i \rangle.$$

Proof: First part follows from the above Lemma.

$$\text{For second part, observe that } \langle w | v_i \rangle = \left\langle \sum_{j=1}^k \alpha_j v_j | v_i \right\rangle$$

$$= \sum_{j=1}^k \alpha_j \langle v_j | v_i \rangle = \alpha_i, \text{ since for } i \neq j, \langle v_j | v_i \rangle = 0$$

$$\text{and } \langle v_i | v_i \rangle = \|v_i\|^2 = 1. \blacksquare$$

Lemma 11.52. If $\{v_1, v_2, \dots, v_k\}$ is an orthonormal subset of V then for any $v \in V$, the vector $v - \sum_{i=1}^k \langle v | v_i \rangle v_i$ is orthogonal to each of

$$v_i, \text{ moreover if } y \in \langle v_1, v_2, \dots, v_k \rangle \text{ then for } z = y - \sum_{i=1}^k \langle y | v_i \rangle v_i,$$

$\|z\| \neq 0$ and for $v_{k+1} = \frac{z}{\|z\|}$, $\{v_1, v_2, \dots, v_k, v_{k+1}\}$ is an orthonormal subset of V .

Proof: Let $j \in \{1, 2, \dots, k\}$, then

$$\langle (v - \sum_{i=1}^k \langle v | v_i \rangle v_i) | v_j \rangle = \langle v | v_j \rangle - \langle v | v_j \rangle = 0 \quad \text{since for } i \neq j, \langle v_i | v_j \rangle = 0 \text{ and } \langle v_i | v_i \rangle = 1.$$

Thus the first part of the lemma is proved.

Now suppose $y \notin \langle v_1, v_2, \dots, v_k \rangle$ and $\|z\| = 0$ where

$$z = y - \sum_{i=1}^k \langle y | v_i \rangle v_i. \quad \text{This in turn implies that } z=0 \text{ or}$$

$$y = \sum_{i=1}^k \langle y | v_i \rangle v_i. \quad \text{So } y \text{ is a linear combination of } v_i\text{'s } (i=1, 2, \dots, k). \text{ In other words } y \in \langle v_1, v_2, \dots, v_k \rangle \text{ and this is against the hypothesis. Hence } \|z\| \neq 0.$$

To complete the proof we must show that $\langle v_{k+1} | v_i \rangle = 0 \quad \forall i=1, 2, \dots, k$ and $\langle v_{k+1} | v_{k+1} \rangle = 1$.

By first part $\langle z | v_i \rangle = 0 \quad \forall i=1, 2, \dots, k$. But $\langle v_{k+1} | v_i \rangle = \langle \frac{z}{\|z\|} | v_i \rangle = \frac{1}{\|z\|} \langle z | v_i \rangle$. So $\langle v_{k+1} | v_i \rangle = 0$ for all $i=1, 2, \dots, k$.

$$\text{Finally } \langle v_{k+1} | v_{k+1} \rangle = \langle \frac{z}{\|z\|} | \frac{z}{\|z\|} \rangle = \frac{1}{\|z\|^2} \langle z | z \rangle = 1.$$

Hence the lemma follows. ■

We are now all set for the famous construction of an orthonormal basis (i.e. an orthonormal spanning set) of a finite dimensional inner product space, starting from a given basis. This method is usually called as *Gram-Schmidt Orthogonalization Process*.

Theorem 11.53. Every finite dimensional inner product space has an orthonormal basis.

Proof: Let V be an inner product space of dimension n . We wish to find an orthonormal set $\{f_1, f_2, \dots, f_n\}$. Since every orthonormal set is L.I. and $\dim V = n$, this will do the job. Let $\{e_1, e_2, \dots, e_n\}$ be a basis of V .

Now to start with, put $v_1 = e_1$. As $e_1 \neq 0$, $\|v_1\| \neq 0$; so taking $f_1 = \frac{v_1}{\|v_1\|}$ we see that $\langle f_1 | f_1 \rangle = 1$. As $e_2 \notin \langle f_1 \rangle$, by Lemma 11.52

there exists a vector $v_2 = e_2 - \langle e_2 | f_1 \rangle f_1$ with $\|v_2\| \neq 0$ and $f_2 = \frac{v_2}{\|v_2\|}$ is such that $\{f_1, f_2\}$ is an orthonormal subset of V . Further $\{f_1, f_2\} \subseteq \langle e_1, e_2 \rangle$. We claim that $e_3 \notin \langle f_1, f_2 \rangle$. Suppose the claim were

not true then we would have found $\lambda_1, \lambda_2 \in F$, at least one of them non-zero, such that $e_3 = \lambda_1 f_1 + \lambda_2 f_2$. This would imply

$$e_3 = \frac{\lambda_1 e_2}{\|v_1\|} + \frac{\lambda_2 e_2}{\|v_2\|} - \lambda_2 \frac{\langle e_1 | f_1 \rangle}{\|v_1\| \|v_2\|} e_1 \text{ which is absurd!}$$

Again by Lemma 11.52, there exists $v_3 = e_3 - \langle e_3 | f_1 \rangle f_1 - \langle e_3 | f_2 \rangle f_2$ with $\|v_3\| \neq 0$ and $f_3 = \frac{v_3}{\|v_3\|}$ is such that $\{f_1, f_2, f_3\}$ is an orthonormal subset of V . Also $\{f_1, f_2, f_3\} \subseteq \langle e_1, e_2, e_3 \rangle$.

Now suppose that we have been able to construct an orthonormal set $\{f_1, f_2, \dots, f_i\} \subseteq \langle e_1, e_2, \dots, e_i \rangle$. As before it can be shown that $e_{i+1} \notin \langle f_1, f_2, \dots, f_i \rangle$. By Lemma 11.52 there exists a vector

$$v_{i+1} = e_{i+1} - \sum_{k=1}^i \langle e_{i+1} | f_k \rangle f_k \text{ with } \|v_{i+1}\| \neq 0 \text{ and } f_{i+1} = \frac{v_{i+1}}{\|v_{i+1}\|}$$

is such that $\{f_1, f_2, \dots, f_{i+1}\}$ is an orthonormal subset of V . Since every orthonormal subset is L.I. and $\dim V = n$, this process terminates till we get an orthonormal set $\{f_1, f_2, \dots, f_n\}$. This is also a basis of V , as $\dim V = n$. Hence V has an orthonormal basis. ■

EXAMPLE 25. Let V be the vector space over $\mathbf{R}$ of polynomials in $\mathbf{R}[x]$ of degree ≤ 3 . Dimension of V over $\mathbf{R}$ is 3. $\{1, x, x^2\}$ is a basis of V . Put $e_1 = 1, e_2 = x$ and $e_3 = x^2$. For $f, g \in \mathbf{R}[x]$, define $\langle f | g \rangle = \int_0^1 f(x) g(x) dx$.

We apply Gram-Schmidt Orthogonalization Process to obtain an orthonormal basis of V over $\mathbf{R}$.

$$\text{Take } v_1 = e_1 \text{ and } f_1 = \frac{e_1}{\|e_1\|} = \frac{e_1}{\sqrt{\int_0^1 e_1 \cdot e_1 dx}} = e_1 = 1.$$

$$\text{Put } v_2 = e_2 - \langle e_2 | f_1 \rangle f_1 = x - \langle x | 1 \rangle = x - \int_0^1 x dx = x - \frac{1}{2}.$$

$$\text{Then } f_2 = \frac{v_2}{\|v_2\|} = \frac{x - \frac{1}{2}}{\sqrt{\int_0^1 (x - \frac{1}{2})^2 dx}} = \sqrt{3}(2x - 1).$$

$$\begin{aligned} \text{Again put } v_3 &= e_3 - \langle e_3 | f_1 \rangle f_1 - \langle e_3 | f_2 \rangle f_2 \\ &= x^2 - \left(\int_0^1 x^2 dx \right) - (6x - 3) \int_0^1 (2x^3 - x^2) dx \\ &= x^2 - x + \frac{1}{6}. \end{aligned}$$

$$\text{Now } \|v_3\|^2 = \int_0^1 (x^2 - x + \frac{1}{6})^2 dx = \frac{1}{180} \Rightarrow \|v_3\| = \frac{1}{6\sqrt{5}}.$$

$$\text{Thus } f_3 = \frac{v_3}{\|v_3\|} = (6x^2 - 6x + 1) \sqrt{5}.$$

Hence $\{1, (2x-1)\sqrt{3}, (5x^2-6x+1)\sqrt{5}\}$ is an orthonormal basis of V .

In the next theorem we show that if W is a subspace of a finite-dimensional inner product space V then $W^\perp$ is a subspace complement to W . So we are justified in calling $W^\perp$, orthogonal complement of W .

Theorem 11.54. For each subspace W of a finite-dimensional inner product space V , $V = W \oplus W^\perp$.

Proof: Since W is a subspace of V and V is an inner product space, W is also an inner product space. By Theorem 11.53, there exists an orthonormal basis $\{w_1, w_2, \dots, w_m\}$ of W .

Let $v \in V$, then $v = \sum_{i=1}^m \langle v | w_i \rangle w_i + x$ where $x = v - \sum_{i=1}^m \langle v | w_i \rangle w_i$.

By Lemma 11.52, $\langle x | w_i \rangle = 0 \quad \forall i = 1, 2, \dots, m$. As any element of W can be written as a linear combination of $w_1, w_2, \dots, w_m$, we get $\langle x | w \rangle = 0$ for all $w \in W$. This gives that $x \in W^\perp$. Hence $v \in W + W^\perp$. Consequently $V = W + W^\perp$. Finally invoking Lemma 11.47 we see that $V = W \oplus W^\perp$. ■

Corollary 11.55. If W is a subspace of a finite-dimensional inner product space V , then $(W^\perp)^\perp = W$.

Proof: Let $x \in W$. For any $v \in W^\perp$, $\langle x | v \rangle = 0 \Rightarrow \langle v | x \rangle = 0 \Rightarrow x \in (W^\perp)^\perp$. This gives that $W \subseteq (W^\perp)^\perp$.

By Theorem 11.54, $V = W \oplus W^\perp$ also $V = W^\perp \oplus (W^\perp)^\perp$. So if $\dim V = n$, $\dim W = m$ then $\dim W^\perp = n - m$ and $\dim (W^\perp)^\perp = n - (n - m) = m$. In this way we get that W is a subspace of $(W^\perp)^\perp$ and $\dim W = \dim (W^\perp)^\perp$. Consequently $W = (W^\perp)^\perp$. ■

WORKED-OUT EXERCISES

Exercise 1. Find if possible all 4-tuples of real numbers a, b, c, d such that for $u = (\alpha_1, \alpha_2)$, $v = (\beta_1, \beta_2) \in \mathbb{R}^2$, $\langle u | v \rangle = a\alpha_1^2 + b\beta_2^2 + c\alpha_1\alpha_2 + d\alpha_2\beta_1$ defines an inner product on $\mathbb{R}^2$.

Solution. Let $\langle u | v \rangle = a\alpha_1^2 + b\beta_2^2 + c\alpha_1\alpha_2 + d\alpha_2\beta_1$, $u = (\alpha_1, \alpha_2)$, and $v = (\beta_1, \beta_2) \in \mathbb{R}^2$. Take $u = (1, 0)$, $v = (0, 0)$, i.e., zero vector of $\mathbb{R}^2$. Then $\langle u | 0 \rangle = 0 \Rightarrow a = 0$. Hence for all $u = (\alpha_1, \alpha_2)$, $v = (\beta_1, \beta_2) \in \mathbb{R}^2$, $\langle u | v \rangle = b\beta_2^2 + c\alpha_1\alpha_2 + d\alpha_2\beta_1$.

Now put $\alpha_1 = \alpha_2 = 0$, $\beta_1 = \beta_2 = 1$, so that u is zero vector of $\mathbb{R}^2$ and $v = (1, 1)$, then $\langle 0 | v \rangle = 0 \Rightarrow b = 0$. Hence $\langle u | v \rangle = c\alpha_1\alpha_2 + d\alpha_2\beta_1$ for all $u = (\alpha_1, \alpha_2)$, $v = (\beta_1, \beta_2) \in \mathbb{R}^2$. This gives that $\langle u | u \rangle = c\alpha_1\alpha_2 + d\alpha_2\alpha_1 \geq 0$, i.e. $(c+d)\alpha_1\alpha_2 \geq 0 \quad \forall \alpha_1, \alpha_2 \in \mathbb{R}$.

In particular $(c+d)\alpha_1^2 \geq 0$ and $(c+d)(-\alpha_1^2) \geq 0$. (Put $\alpha_2 = \alpha_1$ and

$-\alpha_1$, respectively). This forces $(c+d)\alpha_1^2$ to be zero for all $\alpha_1 \in \mathbf{R}$. As a consequence we get that $c = -d$. Thus $\langle u | v \rangle = c(\alpha_1\alpha_2 - \alpha_2\beta_1)$ for all $u = (\alpha_1, \alpha_2)$, $v = (\beta_1, \beta_2) \in \mathbf{R}^2$.

$$\begin{aligned}\text{Finally } \langle u | v \rangle &= \langle v | u \rangle \Rightarrow c(\alpha_1\alpha_2 - \alpha_2\beta_1) = c(\beta_1\beta_2 - \beta_2\alpha_1) \\ &\Rightarrow c\alpha_1(\alpha_2 + \beta_2) = c\beta_1(\alpha_2 + \beta_2) \quad \forall \alpha_1, \alpha_2, \beta_1, \beta_2 \in \mathbf{R} \\ &\Rightarrow c\alpha_1 = c\beta_1 \quad \forall \alpha_1, \beta_1 \in \mathbf{R} \\ &\Rightarrow c = 0.\end{aligned}$$

Hence $a = b = c = d = 0$, in other words the only 4-tuple (a, b, c, d) for which $\langle u | v \rangle = a\alpha_1^2 + b\beta_2^2 + c\alpha_1\alpha_2 + d\alpha_2\beta_1$ with $u = (\alpha_1, \alpha_2)$, $v = (\beta_1, \beta_2) \in \mathbf{R}^2$ might be an inner product, is $(0, 0, 0, 0)$. But this is wrong, since then $\langle u | u \rangle = 0$ without u being zero. Hence such a 4-tuple can never be found.

Exercise 2. (Bessel's Inequality). If $\{w_1, \dots, w_k\}$ is an orthonormal subset of V then prove that $\sum_{i=1}^k |\langle w_i | v \rangle|^2 \leq \|v\|^2$ for all $v \in V$.

Solution. Let $v \in V$. We can write $v = \sum_{i=1}^k \langle v | w_i \rangle w_i + x$ where $x = v - \sum_{i=1}^k \langle v | w_i \rangle w_i$. By Lemma 11.52, $\langle x | w_i \rangle = 0 \quad \forall i = 1, 2, \dots, k$. Putting $w = \sum_{i=1}^k \langle v | w_i \rangle w_i$, we get that $v = w + x$ with

$$\langle w | x \rangle = \left\langle \sum_{i=1}^k \langle v | w_i \rangle w_i \mid x \right\rangle = \sum_{i=1}^k \langle v | w_i \rangle \langle w_i | x \rangle = 0$$

$$\text{Thus } \|v\|^2 = \langle w + x | w + x \rangle = \langle w | w \rangle + \langle x | x \rangle = \|w\|^2 + \|x\|^2 \Rightarrow \|w\|^2 \leq \|v\|^2.$$

$$\begin{aligned}\text{But } \|w\|^2 &= \left\langle \sum_{i=1}^k \langle v | w_i \rangle w_i \mid \sum_{i=1}^k \langle v | w_i \rangle w_i \right\rangle \\ &= \sum_{j=1}^k \sum_{i=1}^k \langle v | w_i \rangle \overline{\langle v | w_j \rangle} \langle w_i | w_j \rangle \\ &= \sum_{i=1}^k |\langle v | w_i \rangle|^2 \text{ as } \langle w_i | w_j \rangle = 0 \text{ for } i \neq j \\ &\quad \text{and } \langle w_i | w_i \rangle = 1 \\ &= \sum_{i=1}^k |\overline{\langle w_i | v \rangle}|^2 = \sum_{i=1}^k |\langle w_i | v \rangle|^2.\end{aligned}$$

$$\text{Hence } \sum_{i=1}^k |\langle w_i | v \rangle|^2 \leq \|v\|^2.$$

Exercise 3. Let W be the set of real-valued functions $y=f(x)$ satisfying $\frac{d^2y}{dx^2}+4y=0$. Prove that (i) W is a subspace of the vector space V over $\mathbf{R}$ of real-valued functions, (ii) dimension of W over $\mathbf{R}$ is 2. If we define $\langle y | z \rangle = \int_0^\pi yz \, dz$ for all $y, z \in W$, find an orthonormal basis of W .

Solution. (i) Let $\mathbf{O}$ be the mapping from $\mathbf{R}$ to $\mathbf{R}$ given by $\mathbf{O}(x)=0 \quad \forall x \in \mathbf{R}$. Then clearly $\mathbf{O} \in W$, so W is non-empty. Let $\lambda_1, \lambda_2 \in \mathbf{R}$ and $y_1, y_2 \in W$ then $\frac{d^2y_1}{dx^2}+4y_1=0$ and $\frac{d^2y_2}{dx^2}+4y_2=0$
 $\Rightarrow \frac{d^2}{dx^2}(\lambda_1 y_1 + \lambda_2 y_2) + 4(\lambda_1 y_1 + \lambda_2 y_2) = \lambda_1 \left(\frac{d^2y_1}{dx^2} + 4y_1 \right) + \lambda_2 \left(\frac{d^2y_2}{dx^2} + 4y_2 \right)$
 $= 0 \Rightarrow \lambda_1 y_1 + \lambda_2 y_2 \in W$. Hence W is a subspace of V .

(ii) By theory of Differential Equations every solution of $\frac{d^2y}{dx^2}+4y=0$ is of the type $c_1 \sin 2x + c_2 \cos 2x$ for some $c_1, c_2 \in \mathbf{R}$. Thus $\{\sin 2x, \cos 2x\}$ span W . Further let $a \sin 2x + b \cos 2x = 0$ for some $a, b \in \mathbf{R}$. Then $\frac{d}{dx}(a \sin 2x + b \cos 2x) = 0 \Rightarrow a \cos 2x - b \sin 2x = 0$. Since $\begin{vmatrix} \sin 2x & \cos 2x \\ \cos 2x & -\sin 2x \end{vmatrix} = -1 \neq 0$, the linear equations $(\sin 2x)a + (\cos 2x)b = 0$ and $(\cos 2x)a + (-\sin 2x)b = 0$ in a and b have a trivial solution, i.e. $a=b=0$.

Hence $a \sin 2x + b \cos 2x = 0 \Rightarrow a=b=0$. Consequently $\{\sin 2x, \cos 2x\}$ is a basis of W over $\mathbf{R}$. Thus $\dim W = 2$.

$$\begin{aligned} \text{Finally we take } w_1 &= \sin 2x, f_1 = \frac{w_1}{\|w_1\|} = \frac{\sin 2x}{\sqrt{\int_0^\pi \sin^2 2x \, dx}} \\ &= \sqrt{\frac{2}{\pi}} \sin 2x. \end{aligned}$$

$$\begin{aligned} \text{Put } w_2 &= \cos 2x - \langle \cos 2x | f_1 \rangle f_1 \\ &= \cos 2x - \left(\int_0^\pi \sqrt{\frac{2}{\pi}} \sin 2x \cos 2x \, dx \right) \sqrt{\frac{2}{\pi}} \sin 2x \\ &= \cos 2x - \frac{2}{\pi} \left(\int_0^\pi \sin 2x \cos 2x \, dx \right) \sin 2x \\ &= \cos 2x. \end{aligned}$$

$$\text{Then } f_2 = \frac{w_2}{\|w_2\|} = \frac{\cos 2x}{\sqrt{\int_0^\pi \cos^2 2x \, dx}} = \sqrt{\frac{2}{\pi}} \cos 2x.$$

Hence $\left\{ \sqrt{\frac{2}{\pi}} \sin 2x, \sqrt{\frac{2}{\pi}} \cos 2x \right\}$ is an orthonormal basis of W over $\mathbf{R}$.

Exercise 4. Prove that two vectors x, y in a Euclidean space are orthogonal if and only if $\|x+y\|^2 = \|x\|^2 + \|y\|^2$. Further show that this result is not necessarily true for unitary spaces.

Solution. Now $\|x+y\|^2 = \langle x+y | x+y \rangle = \langle x | x \rangle + \langle x | y \rangle + \langle y | x \rangle + \langle y | y \rangle = \|x\|^2 + 2\langle x | y \rangle + \|y\|^2$ as $\langle x | y \rangle = \overline{\langle y | x \rangle} = \langle y | x \rangle$, since $\langle y | x \rangle$ is real.

We know that x is orthogonal to y if and only if $\langle x | y \rangle = 0$

$$\Leftrightarrow 2\langle x | y \rangle = 0 \Leftrightarrow \|x+y\|^2 = \|x\|^2 + \|y\|^2.$$

Consider $\mathbf{C}$ with standard inner product. Take $x = (0, i)$, $y = (0, 1)$. Then $\langle x | y \rangle = i \neq 0$; so x is not orthogonal to y .

But $\|x+y\|^2 = \|(0, 1+i)\|^2 = \langle (0, 1+i) | (0, 1+i) \rangle = (1+i)(1-i) = 2$ and $\|x\|^2 = \langle (0, i) | (0, i) \rangle = i(-i) = 1$ and $\|y\|^2 = \langle (0, 1) | (0, 1) \rangle = 1 \cdot 1 = 1$. Hence $\|x+y\|^2 = \|x\|^2 + \|y\|^2$ while x is not orthogonal to y .

PROBLEMS

In the following V is always an inner product space over F .

1. Prove that for all $u, v \in V$, $\|u+v\|^2 + \|u-v\|^2 = 2(\|u\|^2 + \|v\|^2)$.

2. Show that if V is a unitary space then $x, y \in V$ are orthogonal if and only if $\|\alpha x + \beta y\|^2 = \|\alpha x\|^2 + \|\beta y\|^2 \forall \alpha, \beta \in \mathbf{C}$.

3. If x, y are vectors in a Euclidean space such that $\|x\| = \|y\|$ then show that $x+y$ is orthogonal to $x-y$.

4. In $V = \mathbf{R}^3$, show that the Schwarz inequality implies that cosine of an angle is of absolute value at most 1.

5. If W is a subspace of V and if $x \in W$ satisfies $\langle x | w \rangle + \langle w | x \rangle \leq \langle w | w \rangle \forall w \in W$, prove that $\langle x | w \rangle = 0 \forall w \in W$.

6. Let V be the vector space of real valued functions $y = f(x)$ satisfying $\frac{d^3y}{dx^3} - 6 \frac{d^2y}{dx^2} + 11 \frac{dy}{dx} - 6y = 0$.

(a) Prove that V is a 3-dimensional vector space over $\mathbf{R}$.

(b) In V , define $\langle f | g \rangle = \int_{-\infty}^0 fg \, dx$. Prove that this defines an inner product on V . Find an orthonormal basis of V over $\mathbf{R}$.

[Ans. $\{\sqrt{2}e^x, 2(3e^{2x} - 2e^x), (3e^x - 12e^{2x} + 10e^x)\sqrt{6}\}$]

7. If V is finite-dimensional and if $\{w_1, w_2, \dots, w_k\}$ is an ortho-

normal subset of V such that $\sum_{i=1}^k |\langle w_i | v \rangle|^2 = \|v\|^2 \forall v \in V$, then prove that $\{w_1, w_2, \dots, w_k\}$ is a basis of V .

8. Let W_1 and W_2 be two subspaces of a vector space V . If both W_1 and W_2 are inner product spaces, show that $W_1 + W_2$ is also an inner product space.

[Hint. Define $\langle u_1 + v_1 | u_2 + v_2 \rangle = \langle u_1 | u_2 \rangle + \langle v_1 | v_2 \rangle$ for all $u_1, u_2 \in W_1; v_1, v_2 \in W_2$.]

9. If W_1 and W_2 are two subspaces of V , show that

$$(W_1 + W_2)^\perp = W_1^\perp \cap W_2^\perp \text{ and } (W_1 \cap W_2)^\perp = W_1^\perp + W_2^\perp.$$

10. If u and v are vectors in V , then show that $|\langle u | v \rangle| = \|u\| \|v\|$ if and only if u and v are L.D. over F .

(This provides necessary and sufficient condition for equality in Schwarz inequality.)

12

LINEAR TRANSFORMATIONS

To study relationships between different groups and different rings, we introduced in early chapters, the concepts of group homomorphism and ring homomorphism. On the same lines we now introduce the concept of homomorphism among vector spaces.

1. Definition of Linear Transformation

Definition 12.1. Let V_1 and V_2 be two vector spaces over the same field F . A mapping T from V_1 to V_2 is called a *linear transformation* or a *vector space homomorphism* if it satisfies the following properties:

$$(1) \quad T(x+y) = T(x) + T(y)$$

$$(2) \quad T(\alpha x) = \alpha T(x)$$

for all $x, y \in V_1$ and $\alpha \in F$.

Sometimes to emphasize that T is a linear transformation from V_1 to V_2 with both V_1 and V_2 vector spaces over F , we say that T is an F -linear transformation. Further the definitions of F -isomorphisms, epimorphisms etc. are analogous to the corresponding concepts in groups and rings.

EXAMPLE 1. Let $F[x]$ be the ring of all polynomials over a field F . We can regard $F[x]$ as a vector space over F . For any $f = a_0 + a_1x + a_2x^2 + \dots + a_nx^n \in F[x]$, we define its derivative D by $D(f) = a_1 + 2a_2x + \dots + na_nx^{n-1}$. It can be easily seen that for any $f, g \in F[x]$, $\alpha \in F$.

$$D(f+g) = D(f) + D(g), \quad D(\alpha f) = \alpha D(f).$$

So that D is a linear transformation of $F[x]$ into itself. D is not a monomorphism, since $D(2x+1) = 2 = D(2x)$, but $2x+1 \neq 2x$.

EXAMPLE 2. For any field F , consider the mapping $T : F^{(3)} \rightarrow F^{(2)}$ given by $T(\alpha, \beta, \gamma) = (\alpha, \beta) \forall \alpha, \beta, \gamma \in F$. Then for any $x = (\alpha_1, \beta_1, \gamma_1)$, $y = (\alpha_2, \beta_2, \gamma_2) \in F^{(3)}$, $\alpha \in F$.

$$\begin{aligned} T(x+y) &= T(\alpha_1 + \alpha_2, \beta_1 + \beta_2, \gamma_1 + \gamma_2) = (\alpha_1 + \alpha_2, \beta_1 + \beta_2) \\ &= (\alpha_1, \beta_1) + (\alpha_2, \beta_2) = T(\alpha_1, \beta_1, \gamma_1) + T(\alpha_2, \beta_2, \gamma_2) \end{aligned}$$

$$\begin{aligned}\text{and } T(\alpha x) &= T(\alpha\alpha_1, \alpha\beta_1, \alpha\gamma_1) = (\alpha\alpha_1, \alpha\beta_1) = \alpha(\alpha_1, \beta_1) \\ &= \alpha T(\alpha_1, \beta_1, \gamma_1) = \alpha T(x).\end{aligned}$$

So T is linear transformation.

EXAMPLE 3. For a field F , define $T: F^{(3)} \rightarrow F^{(3)}$ by $T(\alpha, \beta) = (\alpha, \beta, 0)$. It can be easily seen that T is a linear transformation.

NOTATION. 'L.T.' will stand for 'Linear Transformation'.

REMARK 1. We show below by examples that the conditions (1) and (2) in the definition of a linear transformation, given above, are independent of each other, in the sense that a mapping T from a vector space U_F into a vector space V_F may satisfy (1) but not (2) or it may satisfy (2) but not (1).

EXAMPLE 4. Let $\mathbf{C}$ be the field of complex numbers. We can regard $\mathbf{C}$ as a vector space over itself. If we define $T: \mathbf{C} \rightarrow \mathbf{C}$ such that for any $z = x + iy: x, y \in \mathbf{R}$, $T(z) = x$, then obviously T satisfies (1). Now $T(2-i) = 2$, $T(5) = 5$, $(2+i)T(2-i) = (2+i)2 = 4+2i$. Since $5 = (2+i)(2-i)$, we see that $T[(2+i)(2-i)] \neq (2+i)T(2-i)$. Hence T does not satisfy the condition (2) of Definition 12.1.

EXAMPLE 5. Define $T: \mathbf{C} \rightarrow \mathbf{R}$, where $\mathbf{C}$ is the field of complex numbers, regarded as a vector space over $\mathbf{R}$, as follows.

$T(z) = \sqrt[3]{x^3 + y^3}$, the real cube root of $x^3 + y^3$, $\forall z = x + iy \in \mathbf{C}$ with $x \in \mathbf{R}, y \in \mathbf{R}$.

Then for any $\alpha \in \mathbf{R}$, $\alpha z = \alpha x + i\alpha y$ gives $T(\alpha z) = \sqrt[3]{\alpha^3 x^3 + \alpha^3 y^3} = \alpha \sqrt[3]{x^3 + y^3} = \alpha T(z)$. Thus T satisfies condition (2) of Definition 12.1. Now $T(2) = 2$, $T(3i) = 3$, $T(2+3i) = \sqrt[3]{8+27} = \sqrt[3]{35}$. Clearly $T(2+3i) \neq T(2) + T(3i)$. Hence T does not satisfy condition (1).

Definition 12.2. A vector space V_1 is said to be isomorphic to a vector space V_2 (both over same field F) if there exists a 1-1 linear transformation of V_1 onto V_2 ; symbolically $V_1 \cong V_2$.

EXAMPLE 6. Let V be an n -dimensional vector space over a field F . Let $\{f_1, f_2, \dots, f_n\}$ be a basis of V . Define $T: V \rightarrow F^{(n)}$ by $T(\sum_{i=1}^n \alpha_i f_i) = (\alpha_1, \alpha_2, \dots, \alpha_n)$. It is left for the readers to check that T is an isomorphism. So that $V \cong F^{(n)}$.

The following theorem provides a very convenient test for a mapping $T: V_1 \rightarrow V_2$ to be a linear transformation.

Theorem 12.3. Let V_1 and V_2 be two vector spaces over the same field F . A mapping $T: V_1 \rightarrow V_2$ is a L.T. if and only if $T(\alpha_1 x_1 + \alpha_2 x_2) = \alpha_1 T(x_1) + \alpha_2 T(x_2)$ for all $x_i \in V_i$ and $\alpha_i \in F (i=1, 2)$.

Proof: Let $T: V_1 \rightarrow V_2$ be a L.T. Then by definition, $T(\alpha_1 x_1) = \alpha_1 T(x_1)$, $T(\alpha_2 x_2) = \alpha_2 T(x_2)$. So that $T(\alpha_1 x_1 + \alpha_2 x_2) = T(\alpha_1 x_1) + T(\alpha_2 x_2) = \alpha_1 T(x_1) + \alpha_2 T(x_2)$.

Conversely let $T: V_1 \rightarrow V_2$ satisfy the hypothesis of the theorem. Consider $x, y \in V$, $\alpha \in F$. By taking $\alpha_1 = 1 = \alpha_2$, $x_1 = x$, $x_2 = y$ we see that $T(x+y) = T(\alpha_1 x_1 + \alpha_2 x_2) = \alpha_1 T(x_1) + \alpha_2 T(x_2) = T(x) + T(y)$.

Again by taking $\alpha_1 = \alpha$, $\alpha_2 = 0$, $x_1 = x$, $x_2 = x$ we get $T(\alpha x) = T(\alpha_1 x_1 + \alpha_2 x_2) = \alpha_1 T(x_1) + \alpha_2 T(x_2) = \alpha T(x)$, since $\alpha_2 T(x_2) = 0$. Hence T is L.T. ■

Theorem 12.4. Let $T: V_1 \rightarrow V_2$ be a linear transformation. If $u_1, u_2, \dots, u_n$ are L.D. vectors in V_1 , then $T(u_1), T(u_2), \dots, T(u_n)$ are L.D. vectors in V_2 .

Proof: Since $u_1, u_2, \dots, u_n$ are linearly dependent in V_1 , there exist $\lambda_1, \lambda_2, \dots, \lambda_n$ in F , not all zero such that $\lambda_1 u_1 + \lambda_2 u_2 + \dots + \lambda_n u_n = 0$. This gives $0 = T(\lambda_1 u_1 + \lambda_2 u_2 + \dots + \lambda_n u_n) = \lambda_1 T(u_1) + \lambda_2 T(u_2) + \dots + \lambda_n T(u_n)$. Hence $T(u_1), T(u_2), \dots, T(u_n)$ are L.D. ■

Let us emphasize that the image of a L.I. set need not be L.I. We leave it to the readers to construct an example concerning the same.

Definition 12.5. (Kernel of a L.T.). For any linear transformation $T: V_1 \rightarrow V_2$, the Kernel of T ($\text{Ker } T$) is defined by $\text{Ker } T = \{v \in V_1 \mid T(v) = 0\}$. $\text{Ker } T$ is also called the null space of T .

Theorem 12.6. Let $T: V_1 \rightarrow V_2$ be a L.T. Then

(i) $\text{Ker } T$ is a subspace of V_1 .

(ii) For any subspace W of V_1 , $T(W)$ is a subspace of V_2 .

Proof: (i) Since $T(0) = 0$ (Prove!), $0 \in \text{Ker } T$, so that $\text{Ker } T \neq \emptyset$.

Let $x, y \in \text{Ker } T$, $\alpha, \beta \in F$.

Then $T(\alpha x + \beta y) = \alpha T(x) + \beta T(y) \dots (\text{Theorem 12.3})$
 $= \alpha 0 + \beta 0 = 0$ gives $\alpha x + \beta y \in \text{Ker } T$. Hence $\text{Ker } T$ is a subspace of V (Theorem 11.6).

(iii) Again $0 \in W \Rightarrow T(0) = 0 \in T(W) \Rightarrow T(W)$ is non-void. Also for any members $T(u), T(v) \in T(W)$ with $u, v \in W$, $\alpha, \beta \in F$, $\alpha T(u) + \beta T(v) = T(\alpha u + \beta v) \in T(W)$ since $\alpha u + \beta v \in W$. Hence $T(W)$ is a subspace of V_2 . ■

REMARK 1. We remind the readers that henceforth we shall be mainly dealing with finite-dimensional vector spaces, unless mentioned otherwise.

REMARK 2. As a particular case of the above theorem, we get that $T(V_1)$ is a subspace of V_2 . So $\dim T(V_1)$ is well-defined.

Definition 12.7. Let $T: V_1 \rightarrow V_2$ be a L.T. The rank of T is defined to be equal to $\dim T(V_1)$, and is denoted by $r(T)$. The nullity of T , denoted by $v(T)$ is defined to be $\dim (\text{Ker } T)$.

REMARK. $T(V_1)$ is also known as the rank space of T .

Lemma 12.8. For any linear transformation $T: V_1 \rightarrow V_2$,
 $r(T) \leq \min (\dim V_1, \dim V_2)$.

Proof : Let $\dim V_1 = n$, $\dim V_2 = m$. Since $T(V_1)$ is a subspace of V_2 , we have $\dim T(V_1) \leq m$ i.e. $r(T) \leq m$.

Now consider any $n+1$ vectors $w_1, w_2, \dots, w_{n+1}$ in $T(V_1)$. For each $i=1, 2, \dots, n+1$ we can find $v_i \in V_1$ such that $T(v_i) = w_i$. As dimension of V_1 is n , the vectors $v_1, v_2, \dots, v_{n+1}$ are L.D. Consequently their images $w_1 = T(v_1), w_2 = T(v_2), \dots, w_{n+1} = T(v_{n+1})$ are L.D. (Theorem 12.4). Hence $T(V_1)$ cannot contain $n+1$ L.I. vectors. So that $\dim T(V_1) \leq n$ i.e. $r(T) \leq n$. All the above arguments now yield that $r(T) \leq \min(m, n)$. ■

Theorem 12.9. (Sylvester Law of Nullity). Let $T: V_1 \rightarrow V_2$ be a L.T., then $\dim V_1 = r(T) + v(T)$.

Proof : Let $\dim V_1 = n$, $r(T) = s$, $v(T) = t$. So that $\dim(\text{Ker } T) = t$. Since every linearly independent subset of V_1 can be extended to a basis of V_1 , we can find a basis $\{u_1, u_2, \dots, u_n\}$ of V_1 such that $\{u_1, u_2, \dots, u_t\}$ is a basis of $\text{Ker } T$. Thus $T(u_i) = 0 \forall i=1, 2, \dots, t$.

Consider any $w \in V_1$, then $w = \alpha_1 u_1 + \alpha_2 u_2 + \dots + \alpha_n u_n$, $\alpha_i \in F$; this gives that $T(w) = \alpha_1 T(u_1) + \alpha_2 T(u_2) + \dots + \alpha_n T(u_n)$

$$= \sum_{i=t+1}^n \alpha_i T(u_i) \text{ since } T(u_i) = 0 \forall 1 \leq i \leq t.$$

This shows that every member of $T(V_1)$ is a linear combination of $T(u_{t+1}), T(u_{t+2}), \dots, T(u_n)$. If we show that these elements are also L.I., we shall get that $\dim T(V_1) = n - t$. Suppose on the contrary there exist $\beta_{t+1}, \beta_{t+2}, \dots, \beta_n \in F$, not all zero such that

$$\beta_{t+1} T(u_{t+1}) + \beta_{t+2} T(u_{t+2}) + \dots + \beta_n T(u_n) = 0.$$

$$\text{This gives that } T\left(\sum_{i=t+1}^n \beta_i u_i\right) = 0 \Rightarrow \sum_{i=t+1}^n \beta_i u_i \in \text{Ker } T$$

$$\Rightarrow \sum_{i=t+1}^n \beta_i u_i = \beta_1 u_1 + \beta_2 u_2 + \dots + \beta_t u_t \text{ for some } \beta_j \in F$$

$$\Rightarrow \beta_1 u_1 + \beta_2 u_2 + \dots + \beta_t u_t - \beta_{t+1} u_{t+1} - \beta_{t+2} u_{t+2} - \dots - \beta_n u_n = 0$$

$\Rightarrow \beta_i = 0 \forall i=1, 2, \dots, n$ as $u_1, u_2, \dots, u_n$ are L.I. We arrive at a contradiction. So we must have $\dim T(V_1) = n - t$ i.e. $n = t + r(T) = v(T) + r(T)$ i.e., $\dim V_1 = v(T) + r(T)$. ■

Theorem 12.10. Let $T: V_1 \rightarrow V_2$ be a L.T. For any subspace $H \subset V_1$, $\dim [T(H)] \geq \dim H - v(T)$.

Proof: Let T_1 be the restriction of T to H . Then T_1 is a L.T. of H into V_2 (Prove!).

$$\text{Thus } \dim H = r(T_1) + v(T_1) = \dim [T_1(H)] + v(T_1) \quad \dots (1)$$

$$\text{Now } T(H) = T_1(H) \Rightarrow \dim [T(H)] = \dim [T_1(H)]. \quad \dots (2)$$

Further since $\text{Ker } T_1 = (\text{Ker } T) \cap H$ (Why?), we have

$$v(T_1) \leq v(T). \quad \dots (3)$$

From (1), (2) and (3) we get

$$\begin{aligned}\dim [T(H)] &= \dim [T_1(H)] = \dim H - v(T_1) \\ &\geq \dim H - v(T).\end{aligned}$$

This proves the theorem. ■

PROBLEMS

(in the following problems, as in the chapter, all vector spaces U, V, W etc. are finite-dimensional over same field.)

1. Show that each of the following mappings is a L.T.

(i) $T: \mathbb{C} \rightarrow \mathbb{C}$ where $\mathbb{C}$ is the vector space of complex numbers over $\mathbb{R}$ with $T(z) = \bar{z}$, where for any $z \in \mathbb{C}$, $\bar{z}$ denotes the complex conjugate of z .

(ii) Let $U = \{(a, b, c) \mid a, b, c \in F, \text{ a field}\}$ be a vector space over F . Define $T: U \rightarrow U$ such that $T(a, b, c) = (0, b, c)$ for all $a, b, c \in F$.

(iii) $T: F[x] \rightarrow F[x]$ where F is a field, such that

$$T[f(x)] = f(x+1) - f(x) \quad \forall f(x) \in F[x].$$

2. Show that a linear transformation $T: U \rightarrow V$ is 1-1 if and only if it maps every linearly independent subset of U into a linearly independent subset of V .

3. For any linear transformation $T: U \rightarrow V$ prove the following:

(i) T is 1-1 if and only if $v(T) = 0$.

(ii) T is onto if and only if $r(T) = \dim V$.

(iii) If $U = V$ then T is 1-1 if and only if T is onto.

4. Let $V_n = \{p(x) \in F[x] \mid \deg p(x) < n\}$.

Let $T: V_n \rightarrow V_n$ be such that $T[p(x)] = p(x+1)$. Show that T is an automorphism of V_n .

5. Let U be a subspace of a vector space V . Define $T: V \rightarrow V/U$ by $T(v) = v + U$, $v \in V$. Show that T is a L.T. of V onto V/U with U as its kernel. Further if f is a L.T. of V onto any vector space W , show that f induces an isomorphism $\bar{f}: V/(Ker f) \rightarrow W$ such that $\bar{f}(v + Ker f) = f(v)$. Deduce that $V/(Ker f) \cong W$. (This is called the *Fundamental Theorem of Homomorphism for vector spaces*).

6. Let $V_n = \{p(x) \in \mathbb{Q}[x] \mid \deg p(x) < n\}$ and $V_{n-1} = \{p(x) \in \mathbb{Q}(x) \mid \deg p(x) < n-1\}$ ($n > 1$). Define $T: V_n \rightarrow V_{n-1}$ by $T[p(x)] = \frac{d}{dx} p(x)$.

Show that T is a L.T. of V_n onto V_{n-1} with $\mathbb{Q}$ as its kernel. Hence prove that $V_n/\mathbb{Q} \cong V_{n-1}$.

7. Show that $F[x]$ has a proper subspace V (infinite-dimensional) such that $F[x] \cong V$.

8. Let U and V be two subspaces of a vector space W . Show that $(U+V)/U \cong V/(V \cap U)$.

Hence deduce that $\dim(U+V) = \dim U + \dim V - \dim(U \cap V)$. (Above isomorphism is called *Second Law of Isomorphism*).

9. Let U and V be two vector spaces of same dimension n , further let $\{x_1, x_2, \dots, x_n\}$ and $\{y_1, y_2, \dots, y_n\}$ be bases of U and V respectively. Prove that the mapping

$T: U \rightarrow V$ such that $T(\sum \alpha_i x_i) = \sum \alpha_i y_i \quad \forall \sum \alpha_i x_i \in U$, is an isomorphism. Deduce that any two vector spaces U and V are isomorphic if and only if they have same dimension.

[Hint : Use Theorem 12.9]

10. Give an example of a mapping $T: U \rightarrow V$ which map linearly dependent vectors of U onto linearly dependent vectors of V , but T is not a L.T.

[Hint : $\mathbb{Q}$ is a vector space over itself. Define $T: \mathbb{Q} \rightarrow \mathbb{Q}$ such that $T(x) = x \quad \forall x (\neq 1) \in \mathbb{Q}$ and $T(1) = 2$].

11. If $V_1, V_2, \dots, V_n$ are vector spaces over the same field F and $T_i: V_i \rightarrow V_{i+1} (i=1, 2, \dots, n-1)$ are L.T.'s, such that

(i) $\text{Ker } T_1 = (0)$.

(ii) $\text{Ker } T_{i+1} = T_i(V_i)$

(iii) $T_{n-1}(V_{n-1}) = V_n$.

Prove that $\sum_{i=1}^n (-1)^i (\dim V_i) = 0$.

12. Let C be the space of continuous functions $f: \mathbb{R} \rightarrow \mathbb{R}$ and define mapping $\phi: C \rightarrow C$ by

$$[\phi(f)](x) = \int_0^x f(t) dt$$

show that ϕ is a linear transformations, such that $\text{Im } \phi$ consists of all continuously differentiable functions which have zero value at zero, whereas $\text{Ker } \phi$ is $\{0\}$.

13. Let V be the vector space over $\mathbb{R}$ of polynomials with real coefficients. Define $\psi: V \rightarrow V$ by

$$\psi\left(\sum_{i=0}^k a_i x^i\right) = \sum_{i=1}^k i a_i x^{i-1}$$

Show that ψ is a linear transformation. Further show that $\text{Ker } \psi = \mathbb{R}$ and $\text{Im } \psi = V$.

14. Let T be a linear transformation from a vector space V to itself such that $\text{Ker } T \subset \text{Im } T$. Prove that dimension of V over F is even.

2. Algebra of Linear Transformations

Let U and V be two vector spaces over the same field F . The set of all linear transformations of U into V is denoted by $\text{Hom}_F(U, V)$

or simply by $\text{Hom}(U, V)$.

Definition 12.11. For any $T_1, T_2 \in \text{Hom}(U, V)$, $\alpha \in F$, the sum $T_1 + T_2$ and scalar multiple αT_1 are mappings of U into V defined by $(T_1 + T_2)(u) = T_1(u) + T_2(u)$ and $(\alpha T_1)(u) = \alpha T_1(u)$ for all $u \in U$.

Lemma 12.12. For any $T_1, T_2 \in \text{Hom}(U, V)$, the sum $T_1 + T_2 \in \text{Hom}(U, V)$.

Proof: Let $x, y \in U$, $\alpha, \beta \in F$. Since T_1 and T_2 are L.T.'s, we have

$$T_1(\alpha x + \beta y) = \alpha T_1(x) + \beta T_1(y)$$

$$T_2(\alpha x + \beta y) = \alpha T_2(x) + \beta T_2(y)$$

Consequently $(T_1 + T_2)(\alpha x + \beta y) = T_1(\alpha x + \beta y) + T_2(\alpha x + \beta y)$

$$= \alpha T_1(x) + \beta T_1(y) + \alpha T_2(x) + \beta T_2(y)$$

$$= \alpha [T_1(x) + T_2(x)] + \beta [T_1(y) + T_2(y)]$$

$$= \alpha [(T_1 + T_2)(x)] + \beta [(T_1 + T_2)(y)]. \blacksquare$$

Lemma 12.13. For $T \in \text{Hom}(U, V)$, $\alpha \in F$, $\alpha T \in \text{Hom}(U, V)$.

Proof: Let $x, y \in U$, $\beta, \gamma \in F$. Then $T(\beta x + \gamma y) = \beta T(x) + \gamma T(y)$; this gives $(\alpha T)(\beta x + \gamma y) = \alpha [\beta T(x)] + \alpha [\gamma T(y)]$.

$$= (\alpha\beta)T(x) + (\alpha\gamma)T(y) = (\beta\alpha)T(x) + (\gamma\alpha)T(y)$$

$$= \beta[(\alpha T)(x)] + \gamma[(\alpha T)(y)].$$

This proves αT is a L.T. and hence $\alpha T \in \text{Hom}(U, V)$. $\blacksquare$

Lemma 12.14. Let U be a finite-dimensional vector space with $B = \{x_1, x_2, \dots, x_n\}$ as its basis. For any vector space V , a mapping $f: B \rightarrow V$ can be extended uniquely to a linear transformation T of U into V . Further two linear transformations T and T' of U into V are equal if and only if $T(x_i) = T'(x_i)$ for all i .

Proof: Since $\{x_1, x_2, \dots, x_n\}$ is a basis of U , any $y \in U$ can be expressed uniquely as $y = \sum_{i=1}^n \alpha_i x_i$, $\alpha_i \in F$.

Define $T: U \rightarrow V$ such that $T(y) = \sum_{i=1}^n \alpha_i f(x_i)$.

T is a L.T. (Verify!). For any fixed $j = 1, 2, \dots, n$ taking $\alpha_i = 0$ for $i \neq j$ and $\alpha_j = 1$, we get $y = x_j$ and thus $T(y) = \sum_{i=1}^n \alpha_i f(x_i) = f(x_j)$.

Hence T extends f . This proves the first part.

For the second part we only show that if T and T' agree on $\{x_1, x_2, \dots, x_n\}$, then they are equal.

To see this consider $x \in U$. Then $x = \sum_{i=1}^n \alpha_i x_i$ for some $\alpha_i \in F$. Since

$$T(x_i) = T'(x_i) \forall i, \text{ we have } T(x) = \sum_{i=1}^n \alpha_i T(x_i) = \sum_{i=1}^n \alpha_i T'(x_i) = T'(x).$$

So that $T = T'$. $\blacksquare$

RAMARK. If V is any vector space with $\{x_1, x_2, \dots, x_n\}$ as its basis and W is any other vector space, the above lemma shows that given any n vectors $y_1, y_2, \dots, y_n \in W$ (need not be L.I.), there exists a linear transformation $T: V \rightarrow W$ such that $T(x_i) = y_i$.

Theorem 12.15. Let U and V be two vector spaces over the same field F . Then $\text{Hom}(U, V)$ is a vector space over F .

Proof: For any $T_1, T_2 \in \text{Hom}(U, V)$, $\alpha \in F$, if we define $(T_1 + T_2)(u) = T_1(u) + T_2(u)$, $(\alpha T_1)(u) = \alpha T_1(u)$ for all $u \in U$, then $T_1 + T_2 \in \text{Hom}(U, V)$ and $\alpha T_1 \in \text{Hom}(U, V)$ (Lemmas 12.12 and 12.13). Thus $\text{Hom}(U, V)$ is closed under addition and scalar multiplication operations. We verify that $\text{Hom}(U, V)$ is a vector space. For this let $T_1, T_2, T_3 \in \text{Hom}(U, V)$, $\alpha, \beta \in F, u \in U$.

(1) $(T_1 + T_2)(u) = T_1(u) + T_2(u) = T_2(u) + T_1(u)$ since addition is commutative in $V \Rightarrow (T_1 + T_2)(u) = (T_2 + T_1)(u) \forall u \in U \Rightarrow T_1 + T_2 = T_2 + T_1$. So that addition is commutative.

$$(2) [(T_1 + T_2) + T_3](u) = (T_1 + T_2)(u) + T_3(u) \\ = [T_1(u) + T_2(u)] + T_3(u)$$

$$= T_1(u) + [T_2(u) + T_3(u)] \text{ since } T_1(u), T_2(u), \\ T_3(u) \in V \text{ and addition is associative in } V. \\ = [T_1 + (T_2 + T_3)](u).$$

Thus $(T_1 + T_2) + T_3 = T_1 + (T_2 + T_3)$.

(3) Define $\bar{0}: U \rightarrow V$ such that for all $u \in U$ $\bar{0}(u) = 0$, the zero vector of V . For any $u_1, u_2 \in U$ we have $\bar{0}(\alpha u_1 + \beta u_2) = 0 = \alpha \bar{0}(u_1) + \beta \bar{0}(u_2)$. So that $\bar{0} \in \text{Hom}(U, V)$.

Now $(T_1 + \bar{0})(u) = T(u) + \bar{0}(u) = T(u) + 0 = T(u)$ gives that $T_1 + \bar{0} = T_1$. Consequently $\bar{0}$ is the additive identity in $\text{Hom}(U, V)$.

(4) Define $T_1': U \rightarrow V$ such that $T_1'(u) = -T_1(u) \forall u \in U$. For any $u_1, u_2 \in U$, $\alpha, \beta \in F$, $T_1'(\alpha u_1 + \beta u_2) = -T_1(\alpha u_1 + \beta u_2) \\ = -[\alpha T_1(u_1) + \beta T_1(u_2)] = \alpha[-T_1(u_1)] + \beta[-T_1(u_2)] \\ = \alpha T_1'(u_1) + \beta T_1'(u_2).$

Consequently T_1' is a L.T., so $T_1' \in \text{Hom}(U, V)$. Now $(T_1 + T_1')(u) = T_1(u) + T_1'(u) = T_1(u) - T_1(u) = 0 = \bar{0}(u)$; we have $T_1 + T_1' = \bar{0}$. This proves $-T_1 = T_1' \in \text{Hom}(U, V)$ i.e. every element of $\text{Hom}(U, V)$ has its additive inverse in $\text{Hom}(U, V)$. Hence $\text{Hom}(U, V)$ is an Abelian group under addition.

$$(5) (i) \text{ Now } [\alpha(T_1 + T_2)](u) = \alpha[(T_1 + T_2)(u)] = \alpha[T_1(u) + T_2(u)] \\ = \alpha T_1(u) + \alpha T_2(u) = (\alpha T_1)(u) + (\alpha T_2)(u) = (\alpha T_1 + \alpha T_2)(u)$$

Therefore $\alpha(T_1 + T_2) = \alpha T_1 + \alpha T_2$.

$$(ii) [(\alpha + \beta)T_1](u) = (\alpha + \beta)T_1(u) = \alpha T_1(u) + \beta T_1(u) \\ = (\alpha T_1)(u) + (\beta T_1)(u) = (\alpha T_1 + \beta T_1)(u).$$

Therefore $(\alpha + \beta)T_1 = \alpha T_1 + \beta T_1$.

$$\begin{aligned} \text{(iii)} \quad [(\alpha\beta)T_1](u) &= (\alpha\beta)T_1(u) = \alpha[\beta T_1(u)] \\ &= \alpha[(\beta T_1)(u)] = [\alpha(\beta T_1)](u). \end{aligned}$$

$$\text{Thus} \quad (\alpha\beta)T_1 = \alpha(\beta T_1).$$

$$\text{(iv)} \quad (1T_1)(u) = 1T_1(u) = T_1(u) \text{ so that } 1T_1 = T_1.$$

Hence $\text{Hom}(U, V)$ is a vector space. ■

Definition 12.16. A ring R is said to be an *algebra* over a field F if the additive group $\langle R, + \rangle$ of R is a vector space over F and for any $x, y \in R, \alpha \in F$, we have $\alpha(xy) = (\alpha x)y = x(\alpha y)$.

Theorem 12.17. For any vector space V_F , $\text{Hom}_F(V, V)$ is an algebra over F .

Proof: Taking $U=V$ in Theorem 12.15, it follows that $\text{Hom}_F(V, V)$ is a vector space over F . For any $T_1, T_2 \in \text{Hom}_F(V, V)$, the composite $T_1T_2: V \rightarrow V$ is such that for any $v_1, v_2 \in V, \alpha, \beta \in F$

$$\begin{aligned} (T_1T_2)(\alpha v_1 + \beta v_2) &= T_1[T_2(\alpha v_1 + \beta v_2)] \\ &= T_1[\alpha T_2(v_1) + \beta T_2(v_2)] = \{\alpha T_1[T_2(v_1)] + \beta T_1[T_2(v_2)]\} \text{ (by Theorem 12.3)} \\ &= \alpha[T_1T_2(v_1)] + \beta[T_1T_2(v_2)]. \end{aligned}$$

This proves that T_1T_2 is a L.T. on V , so that $T_1T_2 \in \text{Hom}_F(V, V)$.

Now for $T_1, T_2, T_3 \in \text{Hom}_F(V, V), v \in V, \alpha \in F$

$$\begin{aligned} [T_1(T_2 + T_3)](v) &= T_1[(T_2 + T_3)(v)] \\ &= T_1[T_2(v) + T_3(v)] \\ &= (T_1T_2)(v) + (T_1T_3)(v) \\ &= (T_1T_2 + T_1T_3)(v). \end{aligned}$$

$$\text{Therefore} \quad T_1(T_2 + T_3) = T_1T_2 + T_1T_3.$$

$$\text{Similarly} \quad (T_2 + T_3)T_1 = T_2T_1 + T_3T_1.$$

Since the resultant composition is an associative operation among all mappings from $V \rightarrow V$, the same holds for all members of $\text{Hom}_F(V, V)$. Thus we get $\text{Hom}_F(V, V)$ is a ring. Now $[T_1(\alpha T_2)](v) = T_1[(\alpha T_2)(v)] = T_1[\alpha T_2(v)] = \alpha[T_1\{T_2(v)\}] = [(\alpha T_1)T_2](v)$. Hence we get $T_1(\alpha T_2) = (\alpha T_1)T_2 = \alpha(T_1T_2)$.

Definition 12.16 now shows that $\text{Hom}_F(V, V)$ is an algebra over F . ■

NOTATION. $\text{Hom}_F(V, V)$ will also be denoted by $A(V)$.

Theorem 12.18. Let U and V be two vector spaces over a field F , of dimensions m and n respectively. Then $\text{Hom}(U, V)$ is a vector space over F of dimension mn .

Proof: Let $\{u_1, u_2, \dots, u_m\}$ and $\{v_1, v_2, \dots, v_n\}$ be bases of U and V respectively. Since every mapping of the basis $\{u_1, u_2, \dots, u_m\}$ into V can be extended to a L.T. of U into V for each $i=1, 2, \dots, m; j=1, 2, \dots, n$ we can find a L.T. $T_{ji}: U \rightarrow V$ such that $T_{ji}(u_k) = 0 \forall k \neq i$ and $T_{ji}(u_i) = v_j$. Then $T_{ji} \in \text{Hom}(U, V) \forall i, j$. If we show that the set $\{T_{ji} \mid i=1, 2, 3, \dots, m; j=1, 2, 3, \dots, n\}$ is a basis of $\text{Hom}(U, V)$

it will follow that $\dim [\text{Hom}(U, V)] = mn$, and that will complete the proof.

Consider $T \in \text{Hom}(U, V)$. Since for each $u_i \in U$, $T(u_i) \in V$ and $\{v_1, v_2, \dots, v_n\}$ is a basis of V , $T(u_i) = \sum_{j=1}^n \alpha_{ji} v_j$ for some $\alpha_{ji} \in F$.

Now $T_{ji}(u_i) = v_j$ and $T_{jk}(u_i) = 0 \forall k \neq i$ give $T(u_i) = \sum_{j=1}^n \alpha_{ji} T_{ji}(u_i)$
 $= \sum_{j=1}^n \left(\sum_{k=1}^m \alpha_{jk} T_{jk} \right) (u_i)$. This gives that $T = \sum_{j,k} \alpha_{jk} T_{jk}$.

Thus we have shown that every $T \in \text{Hom}(U, V)$ is a linear combination of T_{jk} 's. Now suppose for some $\beta_{jk} \in F$, $\sum_{j=1}^n \sum_{k=1}^m \beta_{jk} T_{jk} = 0$. This gives $\left[\sum_{j=1}^n \left(\sum_{k=1}^m \beta_{jk} T_{jk} \right) \right] (u_i) = 0 \forall i = 1, 2, \dots, m$, $\Rightarrow \sum_{j=1}^n \beta_{jk} \left[\sum_{k=1}^m T_{jk}(u_i) \right] = 0 \forall i = 1, 2, \dots, m \Rightarrow \sum_{j=1}^n \beta_{ji} v_j = 0 \forall i = 1, 2, \dots, m \Rightarrow \beta_{ji} = 0 \forall i = 1, 2, \dots, m; j = 1, 2, \dots, n$ since $v_1, v_2, \dots, v_n$ are L. I. over F .

Hence T_{jk} 's are also L.I.
 Consequently $\{T_{jk} \mid j=1, 2, \dots, n; k=1, 2, \dots, m\}$ forms a basis of $\text{Hom}(U, V)$ and hence the result follows. ■

Corollary 12.19. For any vector space V_F of dimension n , $A(V)$ is an algebra of dimension n^2 over F . ■

Theorem 12.20. If $T_1, T_2 \in \text{Hom}(U, V)$, then

(i) $r(\alpha T_1) = r(T_1) \forall \alpha (\neq 0) \in F$.

(ii) $|r(T_1) - r(T_2)| \leq r(T_1 + T_2) \leq r(T_1) + r(T_2)$.

Proof: (i) Now $(\alpha T_1)(U) = \alpha T_1(U) \subseteq T_1(U)$, since $T_1(U)$ is a subspace of V . For same reasons $\alpha^{-1} T_1(U) \subseteq T_1(U)$. Therefore $\alpha[\alpha^{-1} T_1(U)] \subseteq \alpha T_1(U)$, i.e. $T_1(U) \subseteq \alpha T_1(U)$. Hence $T_1(U) = \alpha T_1(U)$. This yields that $\dim T_1(U) = \dim [(\alpha T_1)(U)]$. Hence $r(T_1) = r(\alpha T_1)$. (Definition 12.7)

(ii) For any $u \in U$, $(T_1 + T_2)(u) = T_1(u) + T_2(u)$.

Therefore $(T_1 + T_2)(U) \subseteq T_1(U) + T_2(U) \Rightarrow \dim [(T_1 + T_2)U] \leq \dim [T_1(U) + T_2(U)] \leq \dim [T_1(U)] + \dim [T_2(U)] \Rightarrow$

Now $T_1 = (T_1 + T_2) + (-T_2)$. By what is done above $r(T_1) \leq r(T_1 + T_2) + r(-T_2)$. By (i), taking $\alpha = -1$ we get $r(-T_2) = r(T_2)$.

So that $r(T_1) - r(T_2) \leq r(T_1 + T_2)$

Similarly $r(T_2) - r(T_1) \leq r(T_2 + T_1)$.

From (1), and, (2) it follows that

... (2)

... (3)

$$|r(T_1) - r(T_2)| \leq r(T_1 + T_2) \leq r(T_1) + r(T_2).$$

Hence the result follows. ■

Following examples show that strict inequalities can hold in part (ii) of the above theorem.

EXAMPLE 6. Consider $\mathbf{R}^2$. Let $\{e_1, e_2\}$ be a basis of $\mathbf{R}^2$. Define linear transformations T_1 and T_2 on $\mathbf{R}^2$ as follow :

For any $\alpha_1 e_1 + \alpha_2 e_2$; $\alpha_1, \alpha_2 \in \mathbf{R}$.

$$\begin{aligned} T_1(\alpha_1 e_1 + \alpha_2 e_2) &= (\alpha_1 + 2\alpha_2)e_1 + (3\alpha_1 + 4\alpha_2)e_2 \\ T_2(\alpha_1 e_1 + \alpha_2 e_2) &= (\alpha_1 - 2\alpha_2)e_1 - (3\alpha_1 + 4\alpha_2)e_2. \end{aligned}$$

$$\text{Thus } T_1(e_1) = e_1 + 3e_2, T_1(e_2) = 2e_1 + 4e_2.$$

It can be easily seen that $T_1(e_1)$ and $T_1(e_2)$ are L.I.

$$\text{So that } r(T_1) = 2. \text{ Similarly } r(T_2) = 2.$$

$$\text{Thus } r(T_1) - r(T_2) = 0 \quad \dots (1)$$

Now $(T_1 + T_2)(\alpha_1 e_1 + \alpha_2 e_2) = 2\alpha_1 e_1$ gives $(T_1 + T_2)(\mathbf{R}^2) = \langle e_1 \rangle$, the one dimensional subspace of $\mathbf{R}^2$ generated by e_1 .

$$\text{So that } r(T_1 + T_2) = 1 \quad \dots (2)$$

$$\text{Also } r(T_1) + r(T_2) = 2 + 2 = 4 \quad \dots (3)$$

From (1), (2) and (3) it is immediate that

$$|r(T_1) - r(T_2)| < r(T_1 + T_2) < r(T_1) + r(T_2).$$

We leave it to the readers to prove if T_1 and T_2 are non-zero L.T., then $|r(T_1) - r(T_2)| = r(T_1 + T_2) = r(T_1) + r(T_2)$ can never hold.

Theorem 12.21. Let $T_1 \in \text{Hom}(V, U)$, $T_2 \in \text{Hom}(W, V)$ where U, V, W are vector space over the some field F , then $r(T_1) + r(T_2) - n \leq r(T_1 T_2) \leq \min[r(T_1), r(T_2)]$, where $n = \dim V$.

Proof : Note that $T_1 T_2$ is a L. T. : $W \rightarrow U$ (why ?)

$$\begin{aligned} r(T_1 T_2) &= \dim [T_1 T_2(W)] \\ &= \dim [(T_1 \{T_2(W)\})] \leq \min [\dim T_2(W), \dim U]. \quad (\text{Lemma 12.8}) \\ &\leq \dim T_2(W) = r(T_2) \quad \dots (1) \end{aligned}$$

$$\text{Also } (T_1 T_2)(W) \subseteq T_1(V). \text{ So that } \dim [T_1 T_2(W)] \leq \dim T_1(V) \text{ i.e. } r(T_1 T_2) \leq r(T_1) \quad \dots (2)$$

$$(1) \text{ and } (2) \text{ gives } r(T_1 T_2) \leq \min[r(T_1), r(T_2)]. \quad \dots (3)$$

$$\text{Since } n = \dim V, n = r(T_1) + v(T_1)$$

$$\text{i.e. } v(T_1) = n - r(T_1).$$

$$\text{Let } T'_1 \text{ be the restriction of } T_1 \text{ to } T_2(W). \quad \dots (4)$$

Then $\text{Ker } T'_1 = (\text{Ker } T_1) \cap T_2(W)$ gives

$$v(T'_1) \leq v(T_1). \quad \dots (5)$$

$$\text{From (4) and (5) we obtain } v(T'_1) \leq n - r(T_1). \quad \dots (6)$$

Now $T'_1 [T_2(W)] = T_1 [T_2(W)] = (T_1 T_2)(W)$ gives $r(T'_1) = r(T_1 T_2)$. As $\dim [T'_1(W)] = r(T'_1) + v(T'_1) = r(T_1 T_2) + v(T'_1)$ we get $r(T_2) = r(T_1 T_2) + v(T'_1) \leq r(T_1 T_2) + n - r(T_1)$ by (6).

$$\text{i.e. } r(T_1) + r(T_2) - n \leq r(T_1 T_2). \quad \dots (7)$$

Hence (3) and (7) give

$$r(T_1) + r(T_2) - n \leq r(T_1 T_2) \leq \min [r(T_1), r(T_2)] \blacksquare.$$

Corollary 12.22. If one of T_1 and T_2 is an isomorphism then $r(T_1 T_2) = \min [r(T_1), r(T_2)]$.

Proof: To be definite let T_1 be an isomorphism, then $T_2 = T_1^{-1} (T_1 T_2)$ gives $r(T_2) \leq \min [r(T_1 T_2)] \leq r(T_1 T_2)$. But as $r(T_1 T_2) \leq r(T_2)$ we get $r(T_1 T_2) = r(T_2)$. Similar if T_2 is an isomorphism then $r(T_1 T_2) = r(T_1)$.

Hence the result follows. $\blacksquare$

EXAMPLE 7. Let V be any vector space over a field F , of dimension 5. Let $\{e_1, e_2, e_3, e_4, e_5\}$ be a basis of V_F .

Define $T_1 : V \rightarrow V$; $T_2 : V \rightarrow V$ as under.

$$\text{For any } x = \sum_{i=1}^5 \alpha_i e_i \in V, T_1(x) = \alpha_1 e_1 + \alpha_2 e_2 + \alpha_3 e_3 \text{ and}$$

$$T_2(x) = \alpha_2 e_2 + \alpha_3 e_3 + \alpha_4 e_4.$$

It is immediate that $T_1(V)$ has a basis $\{e_1, e_2, e_3\}$ and $T_2(V)$ has a basis $\{e_2, e_3, e_4\}$.

Thus

$$r(T_1) = r(T_2) = 3 \quad \dots (1)$$

Now $(T_1 T_2)(x) = \alpha_2 e_2 + \alpha_3 e_3$. This gives $(T_1 T_2)(V)$ has a basis $\{e_2, e_3\}$ and so

$$r(T_1 T_2) = 2 \quad \dots (2)$$

From (1) and (2) it is immediate that

$$r(T_1) + r(T_2) - 5 < r(T_1 T_2) < \min [r(T_1), r(T_2)].$$

WORKED-OUT EXERCISES

Exercise 1. Let V be a vector space over a field F and let $W_1, W_2, \dots, W_n$ be subspaces of V such that $V = W_1 \oplus W_2 \oplus \dots \oplus W_n$. For each

$i \in \{1, 2, \dots, n\}$, define $E_i : V \rightarrow W_i$ by $E_i(v) = v_i$ where $v = \sum_{i=1}^n v_i$, $v_i \in W_i$. Show that for $i, j \in \{1, 2, \dots, n\}$

(i) E_i is a linear transformation.

(ii) $E_i(x) = x \forall x \in W_i$.

(iii) $E_i E_j = E_i E_j = 0$ if $i \neq j$, $E_i^2 = E_i$.

(iv) $\sum_{i=1}^n E_i = I$ where I is the identity transformation on V .

Solution. (i) Firstly we show that E_i is well-defined $\forall i \in \{1, 2, \dots, n\}$

$$\text{Let } v = \sum_{i=1}^n v_i = \sum_{i=1}^n v_i' \text{ with } v_i, v_i' \in W_i.$$

Since V is direct sum of $W_1, W_2, \dots, W_n$, we get that $v_i = v_i' \forall i = 1, 2, \dots, n$. Hence $E(v) = v_i = v_i'$. This shows that E_i is well-defined.

Further let $\alpha, \beta \in F$, $u, v \in V$. Then $u = \sum_{i=1}^n u_i$, $v = \sum_{i=1}^n v_i$ with $u_i, v_i \in W_i$ ($1 \leq i \leq n$) $\Rightarrow (\alpha u + \beta v) = \sum_{i=1}^n (\alpha u_i + \beta v_i) \Rightarrow E_i(\alpha u + \beta v) = \alpha u_i + \beta v_i = \alpha E_i(u) + \beta E_i(v) \Rightarrow E_i$ is a L.T. from V to W_i .

(ii) Let $x \in W_i$, then $x \in W$ and so $x = \sum_{i=1}^n u_i$, $u_i \in W_i$. Because V is direct sum of subspaces W_i , we get $u_j = 0 \forall j \neq i$ and $u_i = x$. Then $E_i(x) = u_i = x$.

(iii) Let $i \neq j$. For all $v \in V$, $E_i(v) = v_i$ where $v = \sum_{i=1}^n v_i$, $v_i \in W_i$.

Now $(E_j E_i)(v) = E_j[E_i(v)] = E_j(v_i) = 0$ as $i \neq j$.

Similarly it can be shown that $(E_i E_j)v = 0 \forall v \in V$. Hence $E_i E_j = E_j E_i = 0$.

Further for any $v \in V$, $E_i^2(v) = E_i[E_i(v)] = E_i(v_i)$ where $v = \sum_{i=1}^n v_i$, $v_i \in W_i$. Using (ii) above, we get $E_i^2(v) = v_i = E_i(v)$.

Hence $E_i^2 = E_i$.

(iv) Let $v \in V$. Then $v = \sum_{i=1}^n v_i$, $v_i \in W_i \Rightarrow$

$$\left(\sum_{i=1}^n E_i \right)(v) = \sum_{i=1}^n E_i(v) = \sum_{i=1}^n v_i = v \Rightarrow \sum_{i=1}^n E_i = I.$$

Exercise 2.

Definition 1. A linear transformation $T: V \rightarrow V$ is called a *projection* if $T^2 = T$.

Definition 2. A set of projections on V is called *orthogonal* if the product of any two distinct ones is zero.

Definition 3. A set of orthogonal projections $E_1, E_2, \dots, E_n$ on V is called *supplementary* if $\sum_{i=1}^n E_i = I$ where I is the identity transformation on V .

Show that there is a 1-1 correspondence between direct decompositions of a vector space and finite sets of supplementary orthogonal projections on that vector space.

Solution. Given any direct decomposition $V = W_1 \oplus W_2 \oplus \dots \oplus W_n$, we get the set $\{E_1, E_2, \dots, E_n\}$ of supplementary orthogonal projections, as proved in Exercise 1 above.

Let $\{E_1, E_2, \dots, E_n\}$ be a finite set of supplementary orthogonal projections on V .

Let $v \in V$. Since $\sum_{i=1}^n E_i = I$, where I is the identity transformation on V , we get $v = Iv = (\sum_{i=1}^n E_i)v = \sum_{i=1}^n E_i(v)$. Thus $V \subseteq \sum_{i=1}^n E_i(V)$. Hence $V = \sum_{i=1}^n E_i(V)$. Put $W_i = E_i(V)$ for all $i \in \{1, 2, \dots, n\}$. Then each W_i is a subspace of V such that $V = \sum_{i=1}^n W_i$.

Let $w_i \in W_j \cap (\sum_{j \neq i} W_j)$, then $w_j = \sum_{j \neq i} w_j$ where $w_k \in W_k \forall k \in \{1, 2, 3, \dots, n\}$.

Then $E_i(w_i) = \sum_{j \neq i} E_i(w_j)$. But $w_j \in W_j \Rightarrow w_j = E_j(x)$ for some $x \in V$. Hence $E_i(w_i) = \sum_{j \neq i} E_i E_j(x) = 0$ as $E_i E_j = 0 \forall i \neq j$. Now $w_i \in W_i \Rightarrow w_i = E_i(y)$ for some $y \in V \Rightarrow E_i(w_i) = E_i^2(y) \Rightarrow E_i^2(y) = 0 \Rightarrow E_i(y) = 0$ as $E_i^2 = E_i$. This in turn implies that $w_i = 0$ so the sum $\sum_{i=1}^n W_i$ is direct. Hence we get a direct decomposition of V .

Let P_i be the projection on V associated with the decomposition $V = W_1 \oplus W_2 \oplus \dots \oplus W_n$. Then for each $v \in V$, $v = \sum_{i=1}^n v_i$, $v_i \in W_i$ ($1 \leq i \leq n$). Thus $P_i(v) = v_i$. Now $v_i \in W_i \Rightarrow v_i = E_i(z)$ for some $z \in V \Rightarrow E(v_i) = E_i^2(z) = E_i(z) = v_i$. Hence $P_i(v) = E_i(v) \forall v \in V$. Consequently $P_i = E_i \forall i (1 \leq i \leq n)$. Hence the result follows.

Exercise 3. Prove that two projections E and F have the same rank spaces if and only if $FE = E$ and $EF = F$. Prove that they have the same null spaces if and only if $FE = F$, $EF = E$.

Solution. $E(V) = FE(V) \subseteq F(V)$ as $E(V) \subseteq V$.

Again $F(V) = EF(V) \subseteq E(V)$ as $F(V) \subseteq V$. Hence $E(V) = F(V)$.

Suppose now that $E(V) = F(V)$.

For any $v \in V$, $E(v) = F(w)$ for some $w \in V$ then $FE(v) = FF(w) = F(w) = F(v) \forall v \in V \Rightarrow FE = F$.

Again for any $v \in V$, $F(v) = E(x)$ for some $x \in V \Rightarrow EF(v) = EE(x) = E(x) = E(v) \Rightarrow EF = E$.

For second part we proceed as under.

Suppose $FE = F$, $EF = E$. Now $v \in \text{Ker } E \Rightarrow E(v) = 0 \Rightarrow FE(v) = 0 \Rightarrow F(v) = 0 \Rightarrow v \in \text{Ker } F$. So $\text{Ker } E \subseteq \text{Ker } F$.

Again $x \in \text{Ker } F \Rightarrow F(x) = 0 \Rightarrow EF(x) = 0 \Rightarrow E(x) = 0 \Rightarrow x \in \text{Ker } E$.

Thus $\text{Ker } F \subseteq \text{Ker } E$. Hence $\text{Ker } F = \text{Ker } E$.

Conversely let $\text{Ker } F = \text{Ker } E$. For any $v \in V$, $F(I - F)(v) = F(v) -$

$$F(v) = F(v) - F(v) = 0$$

$$\Rightarrow (I - F)v \in \text{Ker } F = \text{Ker } E$$

$$\Rightarrow E[(I - F)v] = 0 \Rightarrow E(v) = EF(v).$$

$$\text{Hence } E = EF.$$

Similarly $E(I - E)v = 0 \forall v \in V$ gives that $F = FE$.

Exercise 4. Show that if E_1 and E_2 are projections such that $E_1E_2 = E_2E_1$, then E_1E_2 and $E_1 + E_2 - E_1E_2$ are projections.

Solution. $(E_1E_2)^2 = E_1E_2E_1E_2 = E_1E_1E_2E_2$, (as $E_2E_1 = E_1E_2$) $\Rightarrow E_1^2E_2^2 = E_1E_2$. Hence E_1E_2 is a projection.

$$\text{Further note that } (E_1 + E_2)E_1E_2 = E_1^2E_2 + E_2E_1E_1 = 2E_1E_2$$

$$\text{and } E_1E_2(E_1 + E_2) = E_1E_2E_1 + E_1E_2^2 = 2E_1E_2$$

$$\begin{aligned} \text{Then } (E_1 + E_2 - E_1E_2)^2 &= (E_1 + E_2)^2 - 2(E_1 + E_2)E_1E_2 + (E_1E_2)^2 \\ &= E_1^2 + E_2^2 + 2E_1E_2 - 4E_1E_2 + E_1E_2 \\ &= E_1 + E_2 - E_1E_2. \end{aligned}$$

Hence $E_1 + E_2 - E_1E_2$ is a projection.

PROBLEMS

(In the following problems all vector spaces unless otherwise stated, are finite-dimensional).

1. Let $\{e_1, e_2, \dots, e_n\}$ be a basis of V_F and $\lambda_1, \lambda_2, \dots, \lambda_n \in F$ be distinct scalars. If $T, U \in \text{Hom}(V, V)$ are such that $T(e_i) = \lambda_i e_i$ and $TU = UT$, prove that $U(e_i) = \mu_i e_i$ for some scalars $\mu_1, \mu_2, \dots, \mu_n$. What can be said about U if the λ_i 's are not all distinct?

2. If U and T are linear transformations in $A(V)$ such that $UT - TU$ commutes with U , then show that for every positive integer k , $U^k T - TU^k = kU^{k-1}(UT - TU)$.

3. Show that every algebra over F with unity can be embedded in $A(V)$ for some vector space V_F . Generalize this result for algebras without unity.

4. If $\dim V > 1$ prove that $A(V)$ is not commutative.

5. If $T(\neq 0) \in A(V)$ prove that there is an $S \in A(V)$ such that $E = TS \neq 0$ is an idempotent.

6. Let $T \in A(V)$ and let $V_0 = \{v \in V \mid T^k(v) = 0 \text{ for some positive integer } k\}$. Prove that V_0 is a subspace of V and if for any positive integer m , $T^m(v) \in V_0$ then $v \in V_0$.

7. If $T \in A(V)$ is such that each subspace of dimension 1 is mapped into itself, prove that $T = \lambda I$ for some scalar λ , where I is the identity transformation.

[Hint. Let $\{e_1, e_2, \dots, e_n\}$ be a basis of V , then $T(e_i) = \lambda_i e_i \forall i = 1, 2, \dots, n$. Again $T < e_1 + e_2 + \dots + e_n > \subseteq < e_1 + e_1 + \dots + e_n > \Rightarrow \lambda_1 = \lambda_2 = \dots = \lambda_n$.]

8. Find two linear transformations T and U on $\mathbf{R}^2$ such that $TU=0$ but $UT \neq 0$.
9. Give an example of a linear transformation T on $\mathbf{R}^2$ such that $T=0$ but $T \neq 0$.
10. Let V and W be two vector spaces over the same field F and let T be an isomorphism from V onto W . Prove that the mapping $f: A(V) \rightarrow A(W)$ given by $f(U) = TUT^{-1}$ for all $U \in A(V)$, is an isomorphism.
11. If $T \in A(V)$, prove that the set of all linear transformations U on V such that $TU=0$ is a subspace of $A(V)$.
12. $T \in A(V)$ is such that $T - T + I = 0$ where I is the identity transformation on V , show that T is invertible.
13. Assume that the characteristic of a field F is different from 2. Prove that if E_1 and E_2 are projections on a vector space V over F such that $E_1 + E_2$ is itself a projection then E_1 and E_2 are orthogonal.
14. Show that if $E_1, E_2, \dots, E_r$ are projections, on a vector space V such that $E_1(V) = E_2(V) = \dots = E_r(V)$ and $\alpha_1, \alpha_2, \dots, \alpha_r \in F$ satisfy $\sum_{i=1}^r \alpha_i = 1$, then $\sum_{i=1}^r \alpha_i E_i$ is a projection with the same rank space as that of E_i .
15. α, β, γ are linear transformations from a vector space V to itself. Show that $r(\alpha\beta) + r(\beta\gamma) \leq r(\beta) + r(\alpha\gamma)$ (Frobenius Inequality).
16. Let T be a linear transformation from $\mathbf{R}^3$ into $\mathbf{R}^2$, and let U be a linear transformation from $\mathbf{R}^2$ into $\mathbf{R}^3$. Prove that UT is not invertible.
17. Let T be a linear transformation from V to V such that $r(T) = r(T^2)$. Prove that $T(V) \cap \text{Ker } T = \{0\}$.

3. Dual Spaces

Definition 12.23. Let V be a vector space over a field F . Any linear transformation of V into F is called a *linear functional* on V .

Definition 12.24. The vector space $\text{Hom}_F(V, F)$, consisting of all linear functionals on V is called the *dual* of V and is denoted by V^* . A dual space is also called a *conjugate space*.

Since dimension of F as a vector space over itself is one, for any vector space V_F (finite-dimensional), Theorem 12.18 yields that $\dim[\text{Hom}_F(V, F)] = \dim V \cdot \dim F = \dim V$ i.e. $\dim V^* = \dim V$. This relation we prove again, while establishing an important relationship between a vector space and its dual.

Theorem 12.25. If V is an n -dimensional vector space over a field F , having $\{e_1, e_2, \dots, e_n\}$ as its basis, then V^* has a basis $\{e_1^*, e_2^*, \dots, e_n^*\}$,

such that $e_i^*(e_j) = 0$ if $i \neq j$ and $e_i^*(e_i) = 1$. Also $\dim V = \dim V^*$.

Proof: Since every mapping of $\{e_1, e_2, \dots, e_n\}$ into F can be extended to a linear functional on V (Lemma 12.14), for each $i = 1, 2, 3, \dots, n$, we can define linear functional e_i^* on V such that

$$\left. \begin{aligned} e_i^*(e_j) &= 0 \quad \forall j \neq i \\ &= 1 \quad \text{for } j = i \end{aligned} \right\} \quad \dots (1)$$

We show that $B = \{e_1^*, e_2^*, \dots, e_n^*\}$ is a basis of V^* .

Let $f \in V^*$. For each $1 \leq i \leq n$, let $f(e_i) = \alpha_i \in F$. Consider $g = \alpha_1 e_1^* + \alpha_2 e_2^* + \dots + \alpha_n e_n^*$.

For $1 \leq i \leq n$, $g(e_i) = (\alpha_1 e_1^* + \alpha_2 e_2^* + \dots + \alpha_n e_n^*)(e_i) = \alpha_i e_i^*(e_i) = \alpha_i = f(e_i)$, since by (1) $e_j^*(e_i) = 0$ for $j \neq i$.

Thus g and f coincide on the basis $\{e_1, e_2, \dots, e_n\}$ of V . So $f = g$. (Lemma 12.14). This shows that B spans V^* . Suppose for some

$\beta_i \in F$, $1 \leq i \leq n$, $\sum_{i=1}^n \beta_i e_i^* = 0$, then $(\sum_{i=1}^n \beta_i e_i^*)(e_j) = 0 \Rightarrow \beta_j = 0 \quad \forall j = 1,$

$2, \dots, n$. Hence B is also L.I. Consequently B is a basis of V^* . As B has n members we get $\dim V^* = n = \dim V$. This completes the proof. ■

Definition 12.26. Let V_F be a finite-dimensional vector space with $(e_1, e_2, \dots, e_n)$ as its ordered basis. Then the ordered basis $(e_1^*, e_2^*, \dots, e_n^*)$ of V^* such that $e_i^*(e_j) = 0$ for $j \neq i$ and $e_i^*(e_i) = 1$ is called the *dual* of the basis $(e_1, e_2, \dots, e_n)$ of V (or simply a *dual basis* of V^*).

Definition 12.27. For any vector space V_F , the dual V^{**} of V^* is called the *bidual* of V .

Lemma 12.28. If V is a vector space (finite-dimensional) over a field F , then for any $v (\neq 0)$ in V there exists $g \in V^*$ such that $g(v) \neq 0$.

Proof: Since $v \neq 0$, $\{v\}$ is a L.I. subset of V . So it can be extended to an ordered basis, say, $(v = v_1, v_2, \dots, v_n)$ of V . Let $(v_1^*, v_2^*, \dots, v_n^*)$ be the basis of V^* dual to $(v_1, v_2, \dots, v_n)$. Then by definition $v_1^*(v_1) = 1 \neq 0$. Thus taking $g = v_1^*$ we get $g(v) \neq 0$. This completes the proof. ■

Theorem 12.29. (Principle of Duality) If V_F is a finite-dimensional vector space then there exists a canonical isomorphism from V onto V^{**} .

Proof: Define $\sigma : V \rightarrow V^{**}$ as follows.

For any $v \in V$, firstly define $\eta_v : V^* \rightarrow F$ such that $\eta_v(f) = f(v) \quad \forall f \in V^*$. Since for any $f, g \in V^*$, $\alpha, \beta \in F$, $\eta_v(\alpha f + \beta g) = (\alpha f + \beta g)(v) = (\alpha f)(v) + (\beta g)(v) = \alpha[f(v)] + \beta[g(v)] = \alpha[\eta_v(f)] + \beta[\eta_v(g)]$, we get η_v is a linear functional on V^* i.e. $\eta_v \in V^{**}$. Put $\sigma(v) = \eta_v$.

Now for $u, v \in V$; $\alpha, \beta \in F$, $\sigma(\alpha u + \beta v) = \eta_{\alpha u + \beta v} \quad \dots (1)$

Again for any $f \in V^*$, $\eta_{\alpha u + \beta v}(f) = f(\alpha u + \beta v) = \alpha f(u) + \beta f(v)$.
 $= \alpha \eta_u(f) + \beta \eta_v(f) = (\alpha \eta_u + \beta \eta_v)(f) \Rightarrow \eta_{\alpha u + \beta v} = \alpha \eta_u + \beta \eta_v$ (2)

From (1) and (2) we get $\sigma(\alpha u + \beta v) = \alpha \sigma(u) + \beta \sigma(v)$. Hence σ is a L.T.

Now $v \in \text{Ker } \sigma \Rightarrow \sigma(v) = 0 \Rightarrow \eta_v(f) = 0 \forall f \in V^*$
 $\Rightarrow f(v) = 0 \forall f \in V^*$ (3)

If $v \neq 0$ then by Lemma 12.28, there exists $f \in V^*$ such that $f(v) \neq 0$. So that (3) implies $v = 0$. Consequently σ is a monomorphism. That gives $\dim V = \dim \sigma(V)$. However $\dim V = \dim V^* = \dim V^{**}$ (Theorem 12.24), so that $\sigma(V)$ is a subspace of V^{**} such that $\dim \sigma(V) = \dim V^{**}$. This can happen only when $\sigma(V) = V^{**}$. Consequently σ is also onto.

Hence σ is an isomorphism of V onto V^{**} and σ is the desired canonical isomorphism. ■

REMARK. Notice that in the above theorem $\sigma : V \rightarrow V^{**}$ does not depend upon any particular choice of basis of V , that is why it is called *canonical isomorphism*.

Definition 12.30. Let V_F be a vector space and V^* be its dual space. For any subset S of V , $S^\perp = \{f \in V^* \mid f(s) = 0 \forall s \in S\}$ is called the *orthogonal* or *annihilator* of S in V^* . For any subset T of V^* the set $T^\perp = \{v \in V \mid f(v) = 0 \forall f \in T\}$ is called the *orthogonal* or *annihilator* of T in V .

Lemma 12.31. Let V be any vector space over a field F and V^* be its dual. For any subsets S and T of V and V^* respectively, $S^\perp$ and $T^\perp$ are subspaces of V^* and V respectively.

Proof: Since $\bar{0}(s) = 0 \forall s \in S$, where $\bar{0}$ denotes the zero linear functional, we have $\bar{0} \in S^\perp$. So that $S^\perp \neq \emptyset$. Now let $f, g \in S^\perp$, $\alpha, \beta \in F$

Then $(\alpha f + \beta g)(s) = \alpha f(s) + \beta g(s) = 0$ since $f(s) = g(s) = 0$. Thus $\alpha f + \beta g \in S^\perp$. This proves $S^\perp$ is a subspace of V^* .

Again since $f(0) = 0 \forall f \in T$ we get $0 \in T^\perp$. So that $T^\perp \neq \emptyset$. Let $u, v \in T^\perp$, $\alpha, \beta \in F, f \in T$. By definition of $T^\perp$, $f(u) = 0 = f(v)$. Consequently $f(\alpha u + \beta v) = \alpha f(u) + \beta f(v) = \alpha \cdot 0 + \beta \cdot 0 = 0$ and thus $\alpha u + \beta v \in T^\perp$. This proves that $T^\perp$ is a subspace of V . ■

Following lemma is an easy consequence of Definition 12.30 and we leave its proof to the readers.

Lemma 12.32. Let V_F be a vector space and V^* be its dual, then

- (a) For any subsets S_1, S_2 of V , $S_1 \subseteq S_2 \Rightarrow S_2^\perp \subseteq S_1^\perp$.
- (a') For any subsets T_1, T_2 of V^* , $T_1 \subseteq T_2 \Rightarrow T_2^\perp \subseteq T_1^\perp$.
- (b) For any subset S of V , $S \subseteq S^{\perp\perp}$.

(b') For any subset T of V^* , $T \subseteq T^{\perp\perp}$. ■

Theorem 12.33. Let V be a finite-dimensional vector space over a field F . For any subspaces W and W' of V and V^* respectively $\dim W + \dim W^\perp = \dim V$ and $\dim W' + \dim (W')^\perp = \dim V^*$.

Proof: Let $\dim V = n$ and $\dim W = r$. Since every L.I. subset of V can be extended to a basis of V , we can find an ordered basis $(v_1, v_2, \dots, v_n)$ of V such that $(v_1, v_2, \dots, v_r)$ is a basis of W . Let $(v_1^*, v_2^*, \dots, v_n^*)$ be the basis of V^* dual to $(v_1, v_2, \dots, v_n)$. By definition of dual basis, for $1 \leq i \leq r$, $v_j^*(v_i) = 0 \forall j = r+1, r+2, \dots, n$. Consider any $w \in W$, then $w = \alpha_1 v_1 + \alpha_2 v_2 + \dots + \alpha_r v_r$ for some $\alpha_i \in F$ gives $v_j^*(w) = \alpha_1 v_j^*(v_1) + \alpha_2 v_j^*(v_2) + \dots + \alpha_r v_j^*(v_r) = 0 \forall j \geq r+1$.

Hence $v_j^* \in W^\perp \forall j \geq r+1$. Thus the subspace

$$\langle v_{r+1}^*, v_{r+2}^*, \dots, v_n^* \rangle \subseteq W^\perp. \quad \dots (1)$$

Consider any $f \in W^\perp$.

As $f \in V^*$, $f = \beta_1 v_1^* + \beta_2 v_2^* + \dots + \beta_n v_n^*$ for some $\beta_i \in F$. $\dots (2)$

Then for $1 \leq i \leq r$, as $v_i \in W$, $f(v_i) = 0$. Now (2) gives $f(v_i) = \beta_i$ so $\beta_i = 0$ for $1 \leq i \leq r$. Consequently $f = \beta_{r+1} v_{r+1}^* + \beta_{r+2} v_{r+2}^* + \dots + \beta_n v_n^*$ and $f \in \langle v_{r+1}^*, v_{r+2}^*, \dots, v_n^* \rangle$. This yields

$$W^\perp \subseteq \langle v_{r+1}^*, v_{r+2}^*, \dots, v_n^* \rangle \quad \dots (3)$$

From (1) and (3) we get

$$W^\perp = \langle v_{r+1}^*, v_{r+2}^*, \dots, v_n^* \rangle. \quad \dots (4)$$

It is clear from this last equality that $\{v_{r+1}^*, v_{r+2}^*, \dots, v_n^*\}$ is a basis of $W^\perp$ consisting of $n-r$ vectors. So that

$$\dim W^\perp = n - r = \dim V - \dim W.$$

Hence $\dim W + \dim W^\perp = \dim V$.

For the second part, let $\dim W' = t$. Now $\dim V = \dim V^* = n$. So we can find an ordered basis $(f_1, f_2, \dots, f_n)$ of V^* such that $\{f_1, f_2, \dots, f_t\}$ is a basis of W' . We can find a basis $(F_1, F_2, \dots, F_n)$ of V^{**} dual to $(f_1, f_2, \dots, f_n)$.

Now the mapping $\sigma: V \rightarrow V^{**}$ such that for every $v \in V$, $f \in V^*$, $[\sigma(v)](f) = f(v)$ is an isomorphism (Theorem 12.29). So we can find $v_1, v_2, \dots, v_n$ in V such that $\sigma(v_i) = F_i$ for $1 \leq i \leq n$. Then for $1 \leq i, j \leq n$, $j \neq i$ we have $0 = F_i(f_j) = [\sigma(v_i)](f_j) = f_j(v_i)$ and for $i = j$, $1 = F_i(f_i) = f_i(v_i)$. Thus $\{v_1, v_2, \dots, v_n\}$ is a basis for V and $(f_1, f_2, \dots, f_n)$ is a basis of V^* dual to $(v_1, v_2, \dots, v_n)$. Now for $1 \leq i \leq t$, $j \geq t+1$, $f_i(v_j) = 0$ gives $v_j \in (W')^\perp$.

So that $\langle v_{t+1}, v_{t+2}, \dots, v_n \rangle \subseteq (W')^\perp$. $\dots (5)$

Suppose $v \in (W')^\perp$. As $v \in V$ we have $v = \alpha_1 v_1 + \alpha_2 v_2 + \dots + \alpha_n v_n$ for some $\alpha_i \in F$ with $1 \leq i \leq n$.

Then $0 = f_j(v) = \alpha_j$ for $1 \leq j \leq t$ gives $v = \alpha_{t+1} v_{t+1} + \dots + \alpha_n v_n \in \langle v_{t+1}, \dots, v_n \rangle$.

$v_{i+2}, \dots, v_n > .$ So that

$$(W')^+ \subseteq \langle v_{i+1}, v_{i+2}, \dots, v_n \rangle. \quad \dots(6)$$

Hence (5) and (6) give $(W')^+ = \langle v_{i+1}, v_{i+2}, \dots, v_n \rangle$. So that $\dim (W')^+ = n - i = n - \dim W'$.

This proves $\dim W' + \dim (W')^+ = n = \dim V^*$. Hence the theorem follows. ■

Theorem 12.34. If S and T are two subspaces of a finite-dimensional vector space V_F , then

$$(a) \ S^{++} = S.$$

$$(b) \ (S+T)^+ = S^+ \cap T^+.$$

$$(c) \ (S \cap T)^+ = S^+ + T^+.$$

Proof: Let $\dim V = n$, $\dim S = r$, $\dim T = s$.

(a) Now $S \subseteq S^{++}$ (Lemma 12.32).

By Theorem 12.33, $\dim S^+ = n - r$ and $\dim S^{++} = n - \dim S^+ = r = \dim S$. Then S is a subspace of S^{++} such that $\dim S = \dim S^{++}$. This gives $S = S^{++}$.

$$\begin{aligned} (b) \ f \in (S+T)^+ &\Rightarrow f(x) = 0 \ \forall x \in S+T \\ &\Rightarrow f(s) = 0 \text{ and } f(t) = 0 \ \forall s \in S \text{ and } t \in T \text{ since} \\ &\quad S \subseteq S+T \text{ and } T \subseteq S+T \\ &\Rightarrow f \in S^+, f \in T^+ \\ &\Rightarrow f \in S^+ \cap T^+ \end{aligned}$$

$$\text{So } (S+T)^+ \subseteq S^+ \cap T^+.$$

$$\text{Again } g \in S^+ \cap T^+ \Rightarrow g \in S^+, g \in T^+ \quad \dots(1)$$

$$\begin{aligned} &\Rightarrow g(s) = 0, g(t) = 0 \ \forall s \in S, t \in T \\ &\Rightarrow g(s+t) = 0 \ \forall s \in S, t \in T \\ &\Rightarrow g \in (S+T)^+ \end{aligned}$$

$$\text{So } S^+ \cap T^+ \subseteq (S+T)^+ \quad \dots(2)$$

From (1) and (2) we get $(S+T)^+ = S^+ \cap T^+$.

By using Lemma 12.32 and Theorem 12.33 the same results as (a) and (b) can be proved for subspaces S and T of V^* , the dual of V .

(c) Now S^+ and T^+ are subspaces of V^* . So by (b) $(S^+ + T^+)^+ = (S^+)^+ \cap (T^+)^+ = S^{++} \cap T^{++} = S \cap T$.

$$\text{Thus } (S \cap T)^+ = (S^+ + T^+)^{++} = S^+ + T^+. \quad \dots[\text{by (a)}].$$

This proves (c). Hence the theorem follows. ■

Following theorem establishes a duality between subspaces of a vector space V and the subspaces of V^* . This theorem is an immediate consequence of Lemma 12.32 and Theorem 12.34.

Theorem 12.35. Let V_F be a finite-dimensional vector space. Then $W \leftrightarrow W^+$ where W is a subspace of V , is a 1-1 correspondence between subspaces of V and those of V^* such that

$$(a) \ W^{++} = W.$$

$$(b) \ W_1 \subseteq W_2 \text{ if and only if } W_2^+ \subseteq W_1^+.$$

- (c) $(W_1 + W_2)^+ = W_1^+ \cap W_2^+$.
 (d) $W_1^+ + W_2^+ = (W_1 \cap W_2)^+$. ■

PROBLEMS

1. Let $V = F^{(3)}$ and let $e_1 = (1, 0, 0)$, $e_2 = (0, 1, 0)$, $e_3 = (0, 0, 1)$. Determine the basis (e_1^*, e_2^*, e_3^*) of V^* dual to the ordered basis (e_1, e_2, e_3) of V . Let $f_1 = (1, 1, 2)$, $f_2 = (0, 2, 1)$, $f_3 = (0, 0, 5)$ and let F be of characteristic not divisible by 2 and 5. Show that $\{f_1, f_2, f_3\}$ is a basis of V . Find the dual basis $\{f_1^*, f_2^*, f_3^*\}$ and express each f_i^* as a linear combination of e_1^*, e_2^*, e_3^* .

2. Let W be a subspace of a finite-dimensional vector space V_F . For each $f \in V^*$ let $f^* = f|_W$ (the restriction of f to W). Show that $f^* \in W^*$ and $f \mapsto f^*$ is a L.T. of V^* onto W^* whose kernel is $W^\perp$.

3. Write in details, the proof of Theorem 12.35.

4. Let V and W be two finite-dimensional vector spaces over the same field F and $A : V \rightarrow W$, a L.T.

Define $A^* : W^* \rightarrow V^*$ such that $A^*(f) = fA \forall f \in W^*$. Show that A^* is a linear transformation (called the *transpose* of A). If U is another vector space over F and $B : W \rightarrow U$ is a L.T., prove that $(BA)^* = A^*B^*$. Further prove that

(i) $\text{Ker } A^* = [A(V)]^\perp$ and $v(A^*) = v(A)$.

Define $A' : W^* \rightarrow V^*$ and $(A')^* : V \rightarrow W$ as follows.

For any $x \in V$, put $(A')^*(x) = y$ if and only if $(A'f)(x) = f(y) \forall f \in W^*$.

(ii) Using the fact that V is canonically isomorphic to V^{**} prove that $(A')^*$ is a L.T.

(iii) For any L.T. $A : V \rightarrow W$, prove that $A = A^{**}$ and deduce that $A \rightarrow A^*$ is a 1-1 L.T. of $\text{Hom}(V, W)$ onto $\text{Hom}(W^*, V^*)$.

5. If f and g are in V^* such that $f(v) = 0$ implies $g(v) = 0$, prove that $g = \lambda f$ for some $\lambda \in F$.

6. If $V = A \oplus B$ then prove that $V^* = A^\perp \oplus B^\perp$, hence deduce that $V^*/A^\perp \cong B^\perp$.

Hint : Use Theorem 12.34.]

4. Matrices and Linear Transformations

Let U and V be two finite-dimensional vector spaces over the same field F and $T : U \rightarrow V$ be a L.T. Now let $(u_1, u_2, \dots, u_n)$ and $(v_1, v_2, \dots, v_m)$ be fixed ordered bases of U and V respectively. Let us recall that a L.T. from U to V is determined by its effect on a basis of U and conversely any mapping of a basis of U into V can be extended uniquely to a L.T. from U into V . As for $1 \leq i \leq n$, $T(u_i) \in V$, we can express $T(u_i)$ uniquely as

$$T(u_i) = \sum_{j=1}^m \alpha_{ji} v_j ; \alpha_{ji} \in F.$$

Then we obtain the $m \times n$ matrix

$$\left[\begin{array}{cccccc} \alpha_{11} & \alpha_{12} & \dots & \dots & \alpha_{1n} \\ \alpha_{21} & \alpha_{22} & \dots & \dots & \alpha_{2n} \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ \alpha_{m1} & \alpha_{m2} & \dots & \dots & \alpha_{mn} \end{array} \right] = (\alpha_{ji}) \quad \dots (1)$$

as the matrix of $(Tu_1, Tu_2, \dots, Tu_n)$ relative to $(v_1, v_2, \dots, v_m)$. Conversely given a matrix (1), we can define a mapping f from

$\{u_1, u_2, \dots, u_n\}$ into V such that $f(u_i) = \sum_{j=1}^m \alpha_{ji} v_j$. This mapping can be extended uniquely to a L.T. $T' : U \rightarrow V$. Clearly the matrix of $(T'u_1, T'u_2, \dots, T'u_n)$ relative to $(v_1, v_2, \dots, v_m)$ is (1) itself. The matrix (α_{ji}) is called the *matrix of the linear transformation T relative to the ordered bases $(u_1, u_2, \dots, u_n)$ and $(v_1, v_2, \dots, v_m)$* . If further $U=V$ and we take $u_i=v_i$ for $1 \leq i \leq n$ then (α_{ij}) is called the *matrix of T relative to $(u_1, u_2, \dots, u_n)$* .

Lemma 12.36. Let U_F and V_F be two vector spaces of dimensions n and m respectively then $\text{Hom}(U, V)$ is isomorphic to the vector space of all $m \times n$ matrices over F .

Proof: Let $(u_1, u_2, \dots, u_n)$ and $(v_1, v_2, \dots, v_m)$ be ordered bases of U and V respectively. For any $T \in \text{Hom}(U, V)$, we have $T(u_i) = \sum_{j=1}^m \alpha_{ji} v_j$ for unique $\alpha_{ji} \in F$. (As emphasized in the last chapter, there is no distinction between left and right vector spaces over a field) we can write

$$T(u_i) = \sum_{j=1}^m v_j \alpha_{ji} ; \alpha_{ji} \in F \quad \dots (1)$$

Then (α_{ji}) is an $m \times n$ matrix. Let $M_{m,n}(F)$ denote the vector space of all $m \times n$ matrices over F .

Define $\sigma(T) = (\alpha_{ji}) \quad \dots (2)$

Let $T, T' \in \text{Hom}(U, V)$, $\alpha \in F$. Suppose for $1 \leq i \leq n$, $T(u_i) = \sum_{j=1}^m v_j \alpha_{ji}$, $T'(u_i) = \sum_{j=1}^m v_j \beta_{ji}$ for some $\alpha_{ji}, \beta_{ji} \in F$, then

$$(T+T')(u_i) = T(u_i) + T'(u_i) = \sum_{j=1}^m v_j (\alpha_{ji} + \beta_{ji}) \quad \text{and} \quad (\alpha T)(u_i) = \alpha \sum_{j=1}^m v_j \alpha_{ji} =$$

$$\sum \alpha (v_j \alpha_{ji}) = \sum v_j (\alpha \alpha_{ji}).$$

So that $\sigma(T+T') = (\alpha_{ji} + \beta_{ji}) = (\alpha_{ji}) + (\beta_{ji}) = \sigma(T) + \sigma(T')$. $\sigma(\alpha T) =$

$(\alpha\alpha_{ji}) = \alpha(\alpha_{ji}) = \alpha\sigma(T)$. Hence σ is a L.T.

Now $\sigma(T) = \sigma(T') \Rightarrow (\alpha_{ji}) = (\beta_{ji}) \Rightarrow \alpha_{ji} = \beta_{ji} \forall i, j$.

$\Rightarrow T(u_i) = \sum_{j=1}^m v_j \alpha_{ji} = \sum_{j=1}^m v_j \beta_{ji} = T'(u_i)$ for $1 \leq i \leq n$. Hence T and T'

agree on a basis of U . Consequently $T = T'$. (Lemma 12.14). Thus σ is 1-1. The remark before the lemma shows that given any matrix $(\alpha_{ji}) \in M_{m,n}(F)$, there exists a linear transformation T , such that $\sigma(T) = (\alpha_{ji})$. Hence σ is onto. This proves the lemma. ■

Corollary 12.37. $A(U)$ is isomorphic to $M_n(F)$, the ring of all $n \times n$ matrices over F .

Proof : In the Lemma 12.36, taking $U = V$, $u_i = v_i$; $1 \leq i \leq n$ we get that there exists a vector space isomorphism,

$\sigma : A(U) \rightarrow M_n(F)$ such that $\sigma(T) = (\alpha_{ji})$ if and only if

$T(u_i) = \sum_{j=1}^n u_j \alpha_{ji}$, for any $T \in A(U)$.

Let T' be another member of $A(U)$ and $T'(u_i) = \sum_{j=1}^n u_j \beta_{ji}$.

Then $(TT')(u_i) = T \sum_{j=1}^n u_j \beta_{ji} = \sum_{j=1}^n T(u_j) \beta_{ji} = \sum_{j=1}^n \left(\sum_{k=1}^n u_k \alpha_{kj} \beta_{ji} \right)$

$= \sum_{k=1}^n u_k \gamma_{ki}$ where $\gamma_{ki} = \sum_{j=1}^n \alpha_{kj} \beta_{ji}$. Thus $\sigma(TT') = (\gamma_{ki}) = \sigma(T)\sigma(T')$.

Hence σ preserves multiplication. Consequently $A(U) \cong M_n(F)$. ■

Theorem 12.38. Let $T \in \text{Hom}(U, V)$. Let $(u_1, u_2, \dots, u_n)$, $(v_1, v_2, \dots, v_m)$ be ordered bases of U and V respectively and (α_{ij}) be the matrix of T relative to these bases. If $(u'_1, u'_2, \dots, u'_n)$, $(v'_1, v'_2, \dots, v'_m)$ are other bases of U and V respectively, with P and Q as their respective matrices relative to $(u_1, u_2, \dots, u_n)$ and $(v_1, v_2, \dots, v_m)$ respectively, then the matrix of T relative to these bases is given by $(\alpha'_{ij}) = Q^{-1}(\alpha_{ij})P$.

Proof : Since $P = (p_{ij})$ is the matrix of $(u'_1, u'_2, \dots, u'_n)$ relative to $(u_1, u_2, \dots, u_n)$ we have

$$u'_i = \sum_{j=1}^n u_j p_{ji}, \quad 1 \leq i \leq n \quad \dots(1)$$

Similarly if $Q^{-1} = (q_{ij}')$, then

$$v_i = \sum_{j=1}^m v'_j q_{ji}', \quad 1 \leq i \leq m \quad \dots(2)$$

Further

$$T(u_i) = \sum_{j=1}^m v_j \alpha_{ji}, \quad 1 \leq i \leq n \quad \dots(3)$$

Consequently $T(u'_i) = \left(\sum_{j=1}^n u_j p_{ji} \right) = \sum_{j=1}^n T(u_j) p_{ji} \quad \dots [\text{by (1)}]$

$$= \sum_{j=1}^n \left(\sum_{k=1}^m v_k \alpha_{kj} \right) p_{ji} \quad \dots [\text{by (3)}]$$

$$= \sum_{j=1}^n \left[\sum_{k=1}^m \left(\sum_{t=1}^m v_t' q_{tk}' \right) \alpha_{kj} \right] p_{ji} \quad \dots \text{by (2)}$$

$$= \sum_{t=1}^m v_t' \alpha_{ti}'$$

where

$$\alpha_{ti}' = \sum_{k=1}^m \sum_{j=1}^n q_{tk}' \alpha_{kj} p_{ji}.$$

Thus the matrix (α_{ti}') of T relative to the bases $(u_1', u_2', \dots, u_n')$ and $(v_1', v_2', \dots, v_m')$ is $Q^{-1}(\alpha_{kj})P$. ■

Corollary 12.39. If $A = (\alpha_{ij})$ is an $m \times n$ matrix over a field F , then there exist non-singular matrices Q and P such that

$$QAP = \text{diag} \{1, 1, \dots, 1, 0, 0, \dots, 0\}.$$

Proof: Let U_F and V_F be two vector spaces of dimensions n and m respectively. It follows from Lemma 12.36 that there exists $T \in \text{Hom}(U, V)$ such that (α_{ij}) is the matrix of U relative to fixed ordered bases $(u_1, u_2, \dots, u_n)$ and $(v_1, v_2, \dots, v_m)$ of U and V respectively. Let $\rho = r(T)$ then we can find a basis $(u_1', u_2', \dots, u_n')$ of U such that $\text{Ker } T = \langle u_{\rho+1}', u_{\rho+2}', \dots, u_n' \rangle$ (Theorem 12.9), then $T(u_1'), T(u_2'), \dots, T(u_\rho')$ are L.I. in V . So we can find a basis $(v_1', v_2', \dots, v_m')$ of V such that $v_i' = T(u_i')$ for $1 \leq i \leq \rho$.

Now $T(u_i') = 0$ for $1 \leq i \leq \rho$ and $T(u_i') = 0$ for $i \geq \rho + 1$.

Thus the matrix of T with respect to the bases $(u_1', u_2', \dots, u_n')$ and

$(v_1', v_2', \dots, v_m')$ is $\text{diag} \{ \overbrace{1, 1, \dots, 1}^{\rho}, 0, 0, \dots, 0 \}$. However, by the

above theorem $\text{diag} \{ \overbrace{1, 1, \dots, 1}^{\rho}, 0, 0, \dots, 0 \} = (Q')^{-1} A P'$, where P' and Q' are the matrices of $(u_1', u_2', \dots, u_n')$ and $(v_1', v_2', \dots, v_m')$ relative to $(u_1, u_2, \dots, u_n)$ and $(v_1, v_2, \dots, v_m)$ respectively. Taking $Q = (Q')^{-1}$ and $P = P'$, we get the result. ■

Corollary 12.40. If $T \in \text{Hom}(U, U)$ has $A = (\alpha_{ij})$ and $A' = (\alpha_{ij}')$ as its matrices relative to the bases $(u_1, u_2, \dots, u_n)$ and $(u_1', u_2', \dots, u_n')$ of U , then $A' = P^{-1} A P$ where P is the matrix of $(u_1', u_2', \dots, u_n')$ relative to $(u_1, u_2, \dots, u_n)$.

Proof: Taking $U = V$, $u_i = v_i$, $u_i' = v_i'$ for $1 \leq i \leq n$, in the above theorem, we get $P = Q^{-1}$ and $A' = P^{-1} A P$. ■

These two corollaries motivate the following definitions.

Definition 12.41. An $m \times n$ matrix A over a field F is said to be equivalent to an $m \times n$ matrix A' over F if $A' = QAP$ for some non-

singular matrices P and Q over F .

Definition 12.42. An $n \times n$ matrix A over a field F is said to be similar to an $n \times n$ matrix A' over F if $A' = P^{-1}AP$ for some non-singular matrix P over F .

Thus Corollary 12.39 shows that any $m \times n$ matrix over a field F is equivalent to a matrix, $\text{diag} \{1, 1, \dots, 1, 0, 0, \dots, 0\}$.

Definition 12.43. Let $A = (a_{ij})$ be any matrix over a field F . It has n columns, namely $A_j = (a_{1j}, a_{2j}, \dots, a_{mj})$; ($j = 1, 2, \dots, n$). We can regard each of these n columns as a member of $F^{(m)}$. If W is the subspace of $F^{(m)}$ generated by these columns, we say that W is the column space of A . The dimension of W is called the column rank of A . Similar definition can be given for row space and row rank.

Theorem 12.44. Let $T : U_F \rightarrow V_F$ be a L.T. and let $A = (a_{ij})$ be a matrix of T relative to bases $(u_1, u_2, \dots, u_n)$ and $(v_1, v_2, \dots, v_m)$ of U and V respectively. Then $r(T) = \text{column rank of } A$.

$$\text{Proof: Now } T(u_i) = \sum_{j=1}^m v_j \alpha_{ji}, \quad 1 \leq i \leq n. \quad \dots (1)$$

Let column rank of A be s .

Now we can find $i_1, i_2, \dots, i_s$ such that the i_1 th, i_2 th, $\dots, i_s$ th columns of A are L.I. and all the other columns are expressible as linear combination of these columns. Suppose $T(u_{i_1})\beta_1 + T(u_{i_2})\beta_2 + \dots + T(u_{i_s})\beta_s = 0$ for some $\beta_1, \beta_2, \dots, \beta_s \in F$.

$$\text{Then (1) gives } \sum_{j=1}^m v_j \left(\sum_{k=1}^s \alpha_{ji_k} \beta_k \right) = 0$$

$$\Rightarrow \sum_{k=1}^s \alpha_{ji_k} \beta_k = 0, \quad 1 \leq j \leq m \Rightarrow \beta_k = 0 \quad \forall k,$$

since the columns are L.I. Hence $T(u_{i_k}), k = 1, 2, \dots, s$, are L.I.

Consider any $T(u_i)$. Now we can find $\alpha_1, \alpha_2, \dots, \alpha_s \in F$ such that

$$\alpha_{ij} = \sum_{k=1}^s \alpha_{ji_k} \alpha_k, \quad 1 \leq j \leq n. \quad \dots (2)$$

since every column of A is a linear combination of the above mentioned s columns.

(1) and (2) yield that

$$\begin{aligned} T(u_i) &= \sum_{j=1}^m v_j \alpha_{ji} = \sum_{j=1}^m v_j \left(\sum_{k=1}^s \alpha_{ji_k} \alpha_k \right) \\ &= \sum_{k=1}^s \left(\sum_{j=1}^m v_j \alpha_{ji_k} \right) \alpha_k \\ &= \sum_{k=1}^s T(u_{i_k}) \alpha_k. \end{aligned}$$

Now the space $T(U)$ is spanned by $T(u_1), T(u_2), \dots, T(u_n)$. Further $T(u_{i_k}), (k=1, 2, \dots, s)$ are L.I. and all the other $T(u_i)$'s are expressible as linear combinations of these s vectors. Consequently $\{T(u_{i_1}), T(u_{i_2}), \dots, T(u_{i_s})\}$ is a basis of $T(U)$. Hence $r(T) = \dim [T(U)] = s =$ column rank of A . This proves the result. ■

Since the rank of a L.T. does not depend upon the choice of the bases of the given vector spaces and to each $m \times n$ matrix $A = (a_{ij})$ there corresponds a L.T. whose matrix is A for some appropriate vector spaces, we get the following.

Corollary 12.45. *Two $m \times n$ matrices A and B over a field F , have same column rank if and only if they are equivalent.*

The proof is based upon Corollary 12.39 and Theorem 12.44, we leave the details to the readers. ■

PROBLEMS

1. Let U_F and V_F be vector spaces with $(u_1, u_2, \dots, u_n)$ and $(v_1, v_2, \dots, v_m)$ as their respective ordered bases. Let $T: U \rightarrow V$ be a L.T. and $A = (a_{ij})$ be the matrix of T relative to the above bases. Let $(u_1^*, u_2^*, \dots, u_n^*)$ and $(v_1^*, v_2^*, \dots, v_m^*)$ be dual bases of U^* and V^* respectively. Prove the following.

(a) $(v_k^* T)(u_i) = a_{ki} = (\sum a_{kj} u_j^*)(u_i); 1 \leq i \leq n.$

(b) $v_k^* T = \sum a_{kj} u_j^*.$

(c) $T^*(v_k^*) = \sum a_{kj} u_j^*$ (Cf. Problem 4, Section 3)

(d) Matrix of T^* relative to the above bases is A' , the transpose of A .

(e) Column rank of $A' = r(T^*) = r(T) =$ column rank of A .

(f) Row rank $A =$ column rank A' .

(g) Row rank $A =$ column rank A .

2. V is a two-dimensional vector space over a field F . Prove that every element in $A(V)$ satisfies a polynomial of degree 2 over F . Further prove that for all $S, T \in A(V)$, $(ST - TS)^2$ commutes with all elements of $A(V)$.

3. Let $U = F^{(3)}$ and suppose that

$$\begin{pmatrix} 1 & 1 & 2 \\ -1 & 2 & 1 \\ 0 & 1 & 3 \end{pmatrix}$$

is the matrix of $T \in A(V)$ relative to the basis (e_1, e_2, e_3) where $e_1 = (1, 0, 0)$, $e_2 = (0, 1, 0)$, $e_3 = (0, 0, 1)$. Find the matrix of T relative to the basis (f_1, f_2, f_3) where $f_1 = (1, 1, 1)$, $f_2 = (0, 1, 1)$ and $f_3 = (0, 0, 1)$.

$$\left[\text{Ans. } \begin{pmatrix} 4 & 3 & 2 \\ -2 & 0 & -1 \\ 2 & 1 & 2 \end{pmatrix} \right]$$

4. Let V be the vector space of polynomials of degree 3 or less over F . In $A(V)$, define T by $T(\alpha_0 + \alpha_1 x + \alpha_2 x^2 + \alpha_3 x^3) = \alpha_0 + \alpha_1(x+1) + \alpha_2(x+1)^2 + \alpha_3(x+1)^3$. Compute the matrix of T relative to bases (a) $(1, x, x^2, x^3)$, (b) $(1, 1+x, 1+x^2, 1+x^3)$. If the matrix in part (a) is A and that in part (b) is B , find a matrix C so that $B = CAC^{-1}$.

$$\left[\text{Ans. (a)} \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 \\ 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 1 \end{pmatrix} \text{ (b)} \begin{pmatrix} 1 & 1 & -1 & -5 \\ 0 & 1 & 2 & 3 \\ 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 1 \end{pmatrix} \right]$$

$$C = \begin{pmatrix} 1 & -1 & -1 & -1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

13

FIELDS

While discussing theory of rings we introduced the concept of a field, as a commutative ring with unity $1 \neq 0$ in which each non-zero element is a unit, and reader must have noticed how beautiful is the theory of vector spaces over a field. The concept of field has a central place in algebra. It has wide applications in Linear Algebra and in the Theory of Equations, which deals with the study of roots of polynomials. Here we do not discuss all aspects of theory of fields and shall be mainly concerned with the development of that part of the theory of fields which has direct applications to the Theory of Equations.

1. Subfields and Prime Fields

Definition 13.1. (Subfield).

A non-void subset S of a field F is said to be a *subfield* of F if (i) $a \in S, b \in S \Rightarrow a+b \in S, ab \in S$ (ii) S is a field under the induced addition and multiplication compositions.

REMARK. If a subfield S of a field F is not equal to F , we shall say that S is a *proper subfield* of F . If S is a subfield of a field F , then S is a subgroup of F under addition and the set of all non zero elements of S form a subgroup of the multiplicative group of F . Consequently for any $a, b \in S, a-b \in S$, further if $b \neq 0$ then $b^{-1} \in S$ and $ab^{-1} \in S$. Since identity of a group belongs to each subgroup, the zero and the unity of F are both members of S . Thus S has at least two elements namely 0 and 1.

On the familiar lines (as for subgroups and subrings) the following important characterization of a subfield can be established.

Theorem 13.2. Any subset S of a field F , containing at least two elements, is a subfield of F if and only if

$$(i) a \in S, b \in S \Rightarrow a-b \in S$$

and

$$(ii) a \in S, b \in S, b \neq 0 \Rightarrow ab^{-1} \in S. \blacksquare$$

Theorem 13.3. *Intersection of any non-void family of subfields of a field F is a subfield of F .*

Proof : Let $[S_i]_{i \in I}$ be a non-empty family of subfields of F and $S = \bigcap_{i \in I} S_i$. Since every subfield of F contains 0 and 1, we get $0, 1 \in S_i \forall i$. Consequently $0, 1 \in S$; so S contains at least two elements. Let $a, b \in S$ then $a \in S_i, b \in S_i \forall i \Rightarrow a - b \in S_i \forall i$, (Theorem 13.2).

When $b \neq 0$, then $ab^{-1} \in S_i \forall i \in I$ gives $ab^{-1} \in \bigcap_{i \in I} S_i = S$. Hence S is a subfield of F . This proves the theorem. ■

Let P be the intersection of all subfields of a field F . Then P is a subfield of F and every subfield of F contains P . So that no subfield of F is contained properly in P . Hence P has no proper subfield. This leads us to define of concept of prime subfield.

Definition. 13.4. A field P is said to be a *prime field* if it has no proper subfields.

EXAMPLE 1. Let p be any prime integer. Then $\langle p \rangle$ is a maximal ideal of $\mathbb{Z}$ and $\mathbb{Z}/\langle p \rangle$ is a field having p elements. If S is any subfield of $\mathbb{Z}/\langle p \rangle$ having at least two elements, then as a group of prime order has no proper subgroups, $S = \mathbb{Z}/\langle p \rangle$. This proves $\mathbb{Z}/\langle p \rangle$ is a prime field.

EXAMPLE 2. Consider $\mathbb{Q}$. Let X be a subfield of $\mathbb{Q}$, then $1 \in X$. Now let $\frac{m}{n} \in \mathbb{Q}, m, n \in \mathbb{Z}, n \neq 0$. Now $m = m \cdot 1 \in X$ and $n = n \cdot 1 \in X$.

But $n \neq 0$ so $n = n \cdot 1 \neq 0 \in X$. This gives $\frac{m}{n} \in X$. Hence $\mathbb{Q} = X$. This shows that $\mathbb{Q}$ is a prime field.

Following theorem shows that $\mathbb{Q}$ and for any prime $p, \mathbb{Z}/\langle p \rangle$ are essentially the only prime fields.

Theorem 13.5. *Any prime field is either isomorphic to the field of rational numbers or to the field of integers modulo some prime number.*

Proof : Let P be a prime field and e be the unity of P . Define $f: \mathbb{Z} \rightarrow P$ by $f(n) = ne \forall n \in \mathbb{Z}$. f is a ring homomorphism and $\text{Ker } f$ is an ideal of $\mathbb{Z}$. We can find a non-negative integer q such that $\text{Ker } f = \langle q \rangle$. Following cases arise:

Case 1 $q = 0$, then f is a monomorphism and $\mathbb{Z} \cong f(\mathbb{Z}) \subseteq P$. Now P contains field of quotients Q' of $f(\mathbb{Z})$, (Theorem 8.25). As P is prime field, we conclude that $P = Q'$. However $\mathbb{Q}$ is the field of quotients of $\mathbb{Z}$. But we know that any two isomorphic domains have isomorphic fields of quotients, we get $\mathbb{Q} \cong Q'$ i.e. $P \cong \mathbb{Q}$.

Case II. $q \neq 0$, then $q > 0$. If $q=1$, we shall get $1e=0$ i.e., $e=0$ that is a contradiction. Hence $q > 1$. We claim that q is a prime number. Let, on the contrary $q=ab$ for some integers a, b , both greater than one. Then $0=qe=(ab)e=(ae)(be)$ gives $ae=0$ or $be=0$. So $a \in \langle q \rangle$ or $b \in \langle q \rangle$ and that gives $q \mid a'$ or $q \mid b$. But this contradicts the fact that a and b are both less than q . Thus q is a prime number. Consequently $\mathbf{Z}/\langle q \rangle$ is a field. Now by the Fundamental Theorem of Homomorphism $\mathbf{Z}/\text{Ker } f \cong f(\mathbf{Z})$ i.e. $\mathbf{Z}/\langle q \rangle \cong f(\mathbf{Z})$.

Thus $f(\mathbf{Z})$ is a subfield of P . That gives $f(\mathbf{Z})=P$ as P is a prime field. This in turn implies that $\mathbf{Z}/\langle q \rangle \cong P$. Hence the result follows. ■

REMARK. In the proof of above theorem, in case (I) the isomorphism $\mathbf{Q} \cong P$ is given by the mapping $x = \frac{n}{m} \rightarrow \frac{ne}{me} \forall x = \frac{n}{m} \in \mathbf{Q}$ and in case (II) the isomorphism $\mathbf{Z}/\langle q \rangle \cong P$ is given by the mapping $\bar{n} \rightarrow ne \forall \bar{n} = n + \langle q \rangle \in \mathbf{Z}/\langle q \rangle$.

Definition 13.6 The intersection P of all subfields of a field F is called the *prime subfield* of F .

Definition 13.7. Let D be an integral domain with unity $e \neq 0$. Then D is said to be of *characteristic zero* if $ne \neq 0$ for every positive integer n . D is said to be of *finite characteristic* if $ne=0$ for some positive integer n . In the latter case if s is the smallest positive integer such that $se=0$ then D is said to be of *characteristic s* .

For example $\mathbf{Q}$ is of characteristic zero and for any positive prime p , $\mathbf{Z}/\langle p \rangle$ is of characteristic p .

Theorem 13.8. The characteristic of a domain D with unity is either zero or a prime number.

Proof : Let D have finite characteristic p . Then $p > 1$ as $1e=e \neq 0$. Suppose $p=ab$ for some integers $a, b > 1$. Then $1 < a, b < p$. Further $0=pe=(ae)(be) \Rightarrow ae=0$ or $be=0$. This contradicts the fact that p is the smallest positive integer with the property that $pe=0$. Hence p is a prime number. This proves the theorem. ■

Worked out Exercises

Exercise 1. Show that the only homomorphisms of $\mathbf{R}$ into itself are zero homomorphism or identity homomorphism.

Solution. Let $f : \mathbf{R} \rightarrow \mathbf{R}$ be a homomorphism.

As $\text{Ker } f$ is an ideal of $\mathbf{R}$ and $\mathbf{R}$ is a field, we get

$$\text{Ker } f = \{0\} \text{ or } \text{Ker } f = \mathbf{R}.$$

In second case $f(x)=0 \forall x \in \mathbf{R} \Rightarrow f$ is zero homomorphism.

In first case f is 1-1.

Since $\text{Ker } f \neq \mathbf{R}$, there exists $r \in \mathbf{R}$ such that $r \notin \text{Ker } f$.
In other words $f(r) \neq 0$.

$$\begin{aligned} \text{But } f(r) &= f(r1) = f(r) \cdot f(1) \\ \Rightarrow f(r) \{1 - f(1)\} &= 0 \\ \Rightarrow f(1) &= 1. \end{aligned}$$

$$\begin{aligned} \text{Let } m \text{ be integer } > 0 \text{ then } f(m) &= f(1 + 1 + 1 + \dots m \text{ times}) \\ &= f(1) + f(1) + f(1) + \dots m \text{ times} \\ &= 1 + 1 + 1 + \dots m \text{ times} \\ &= m. \end{aligned}$$

$$\begin{aligned} \text{When } m=0 \text{ then } f(m) &= f(0) = 0 \text{ as } f \text{ is homomorphism} \\ \Rightarrow f(m) &= m. \end{aligned}$$

$$\begin{aligned} \text{Let } m \text{ be negative integer. Put } m &= -n \text{ where } n > 0. \\ \text{Then } f(m) &= f(-n) = -f(n) \text{ as } f \text{ is homomorphism} \\ &= -n \text{ as } n \text{ is a positive integer} \\ &= m \end{aligned}$$

$$\text{Let } r \text{ be any rational number then } r = \frac{p}{q}, \quad p, q \in \mathbf{Z}, q > 0$$

$$\begin{aligned} &qr = p \\ \Rightarrow f(qr) &= f(p) = p \text{ as proved above.} \\ \Rightarrow f(r + r + \dots q \text{ times}) &= p \\ \Rightarrow f(r) + f(r) + \dots q \text{ times} &= p \\ \Rightarrow q f(r) = p \Rightarrow f(r) &= \frac{p}{q} = r \end{aligned}$$

$$\text{Let } \alpha \geq 0 \text{ then } \alpha = \epsilon^2, \epsilon \geq 0$$

$$\begin{aligned} \text{So } f(\alpha) &= f(\epsilon^2) = [f(\epsilon)]^2 \text{ as } f \text{ is homomorphism} \\ \Rightarrow f(\alpha) &\geq 0 \text{ since } f(\epsilon) \in \mathbf{R}. \end{aligned}$$

$$\begin{aligned} \text{Further } \alpha \geq \beta &\Rightarrow \alpha - \beta \geq 0 \Rightarrow f(\alpha - \beta) \geq 0 \Rightarrow f(\alpha) - f(\beta) \geq 0 \\ \Rightarrow f(\alpha) &\geq f(\beta). \end{aligned}$$

Let $\alpha > \beta$ be such that $f(\alpha) = f(\beta)$. Then as f is 1-1, we would have $\alpha = \beta$, a contradiction. Hence $\alpha > \beta \Rightarrow f(\alpha) > f(\beta)$.

Suppose $\epsilon > 0$, there exists a rational number s such that $0 < s < \epsilon$.

$$\text{Chose } \delta = s \text{ then for } |x - a| < \delta = s$$

$$\begin{aligned} \Rightarrow -s &< x - a < s \\ \Rightarrow -f(s) &< f(x - a) < f(s) \\ \Rightarrow -f(s) &< f(x) - f(a) < f(s) \\ \Rightarrow |f(x) - f(a)| &< f(s) \\ \Rightarrow |f(x) - f(a)| &< s \text{ as } f(s) = s, s \text{ being a rational} \\ \Rightarrow |f(x) - f(a)| &< \epsilon. \end{aligned}$$

number

Hence $f(x)$ is a continuous real valued function.
 Finally let x be any real number.

Then $x = \lim_{n \rightarrow \infty} \left\{ \frac{p_n}{q_n} \right\}$, where $\left\{ \frac{p_n}{q_n} \right\}$ is cauchy sequence of rational numbers converging to x .

Because of continuity of f , we get

$$f(x) = \lim_{n \rightarrow \infty} f\left(\frac{p_n}{q_n}\right) = \lim_{n \rightarrow \infty} \frac{p_n}{q_n} = x.$$

Consequently f is identity homomorphism.

Exercise 2 Show that only continuous homomorphisms of $\mathbf{C}$ into itself are zero homomorphism or conjugation homomorphism or identity homomorphism.

Solution. Since $\mathbf{C}$ is a field, as in above exercise we get that if f not zero homomorphism, then

$$f(r) = r \text{ for every rational number } r.$$

Let x be any real numb. r .

There exists a cauchy sequence $\left\{ \frac{p_n}{q_n} \right\}$ of rational number such that

$$x = \lim_{n \rightarrow \infty} \left\{ \frac{p_n}{q_n} \right\}$$

As f is continuous, by hypothesis

$$f(x) = \lim_{n \rightarrow \infty} f\left(\frac{p_n}{q_n}\right) = \lim_{n \rightarrow \infty} \frac{p_n}{q_n} = x$$

Consider

$$\Rightarrow i = \sqrt{-1}, i^2 = -1 \Rightarrow f(i^2) = f(-1) = -1$$

$$\Rightarrow f(i) = i \quad \text{or} \quad -i.$$

In first case for any $z = \alpha + i\beta \in \mathbf{C}$, ($\alpha, \beta \in \mathbf{R}$)

$$\begin{aligned} f(z) &= f(\alpha + i\beta) = f(\alpha) + f(i)f(\beta) \\ &= \alpha + i\beta \\ &= z. \end{aligned}$$

Hence f is identity homomorphism.

In second case for any $z = \alpha + i\beta \in \mathbf{C}$ ($\alpha, \beta \in \mathbf{R}$)

$$\begin{aligned} f(z) &= f(\alpha + i\beta) = f(\alpha) + f(i)f(\beta) \\ &= \alpha - i\beta. \end{aligned}$$

Consequently f is conjugation homomorphism.

PROBLEMS

1. Prove the following sets are subfields of the field $\mathbf{R}$.

(i) $\{a + b\sqrt{2} \mid a, b \in \mathbf{Q}\}$

$$(ii) \{a + \sqrt[3]{5}b + \sqrt[3]{5}c \mid a, b, c \in \mathbb{Q}\}$$

$$(iii) \{a + \sqrt{2}b + \sqrt{3}c + \sqrt{6}d \mid a, b, c, d \in \mathbb{Q}\}$$

2. Prove that the characteristic of a field is equal to the characteristic of its subfields.

3. Prove that the characteristic of a field F is zero or a prime p according as its prime subfield P is isomorphic to $\mathbb{Q}$ or to $\mathbb{Z}/\langle p \rangle$.

4. Prove or disprove that if D is a PID and p is any irreducible element of D then the characteristic of $D/\langle p \rangle$ is finite.

$$\left[\text{Hint : Consider the field } = \frac{\mathbb{Q}[x]}{\langle x \rangle} \right]$$

5. Prove that if for some non-zero $a \in F$ (F is a field), $na = 0$ for some integer n , then $nb = 0$ for all $b \in F$. Deduce that if for a non-zero element x , there exists a positive integer n such that $nx = 0$, then F is of finite characteristic.

2. Extension of Fields

Definition 13.9. A *Field Extension* of a field F is a pair (K, σ) , where K is a field and σ is a monomorphism of F into K .

Suppose E is a field and F is a subfield of E . The injection map $i: F \rightarrow E$ defined by $i(x) = x$ for all $x \in F$ is a monomorphism. Consequently (E, i) is a field extension of F . In such a situation when i is a trivial type of mapping, we do not mention i and simply say that E is a field extension of F . In general case of a field extension (K, σ) , K has a subfield $\sigma(F)$ isomorphic to F . As emphasized earlier we can identify any two isomorphic algebraic systems. Here also we shall normally identify F with $\sigma(F)$ and thus treat F as a subfield of K , then we shall simply say that K is a field extension of F . Henceforth F will be a field and K will be a field extension of F .

Now, as for any $\alpha \in F$, $x \in K$, $\alpha x \in K$, we get an external composition $F \times K \rightarrow K$ given by $(\alpha, x) \rightarrow \alpha x$. One can immediately see that the additive group $\langle K, + \rangle$ becomes a vector space over F relative to the external composition in K with respect to F defined above. Thus K must have a basis and dimension over F .

Definition 13.10. The dimension of K as a vector space over F is called the *degree* of K .

NOTATION. $[K : F]$ will denote the degree of K over F .

EXAMPLE 3. $\mathbb{Q}[\sqrt{2}] = \{a + b\sqrt{2} \mid a, b \in \mathbb{Q}\}$ is field extension of $\mathbb{Q}$. The subset $\{1, \sqrt{2}\}$ forms a basis of $\mathbb{Q}[\sqrt{2}]$ over $\mathbb{Q}$. Consequently $[\mathbb{Q}[\sqrt{2}] : \mathbb{Q}] = 2$.

EXAMPLE 4. Consider an indeterminate x over a field F . Let K be the field of quotients of $F[x]$. Then, as we know that for any

$$\alpha_0, \alpha_1, \alpha_2, \dots, \alpha_n \in F$$

$$\alpha_0 1 + \alpha_1 x + \alpha_2 x^2 + \dots + \alpha_n x^n = 0$$

implies $\alpha_i = 0 \forall i$, it follows that $\{1, x, x^2, x^3, \dots, x^n, \dots\}$ is an infinite subset of K which is linearly independent over F . Consequently $[K : F]$ is infinite.

Definition 13.11. K is said to be a *finite or infinite extension* of F according as the degree of K over F is finite or infinite.

Thus in Example 3, $\mathbb{Q}[\sqrt{2}]$ is a finite extension of $\mathbb{Q}$ and in Example 4, K is an infinite extension of F . Following theorem about degrees of field extensions is of great significance.

Theorem 13.12. If K is a finite field extension of F and L is a finite field extension of K then L is a finite field extension of F and $[L : F] = [L : K][K : F]$.

Proof: Let $[K : F] = n$ and $[L : K] = m$. Then we can find a basis $\{x_1, x_2, \dots, x_n\}$ of K over F and a basis $\{y_1, y_2, y_3, \dots, y_m\}$ of L over K . Consider the elements $x_i y_j$; $i = 1, 2, \dots, n$; $j = 1, 2, \dots, m$. If we can show that these mn elements of L are linearly independent over F and they generate L as a vector space over F , it will give that $[L : F] = mn = [L : K][K : F]$ and the theorem will follow. For their linear independence let $\sum_{i,j} \alpha_{ij} x_i y_j = 0$ for some $\alpha_{ij} \in F$. That gives

$$\sum_{j=1}^m \left(\sum_{i=1}^n \alpha_{ij} x_i \right) y_j = 0. \text{ Since } y_1, y_2, \dots, y_m \text{ are linearly independent over}$$

K and $\sum_{i=1}^n \alpha_{ij} x_i \in K$ for every j , we get $\sum_{i=1}^n \alpha_{ij} x_i = 0 \forall j = 1, 2, \dots, m$.

This gives $\alpha_{ij} = 0 \forall i, j$ as $x_1, x_2, \dots, x_n$ are linearly independent over F . Hence the elements $x_i y_j$ are linearly independent over F . Consider any $x \in L$. Then

$$x = \sum_{j=1}^m a_j y_j \text{ for some } a_j \in K, j = 1, 2, \dots, m.$$

as $y_1, y_2, \dots, y_m$ is a basis of L over K . Again for each j

$$a_j = \sum_{i=1}^n \alpha_{ij} x_i \text{ for some } \alpha_{ij} \in F, \text{ as } x_1, x_2, \dots, x_n \text{ is a basis of } K \text{ over } F. \text{ Consequently}$$

$$x = \sum_{j=1}^m \sum_{i=1}^n \alpha_{ij} x_i y_j \text{ with } \alpha_{ij} \in F \forall i, j.$$

Hence the mn elements $x_i y_j$, $i = 1, 2, \dots, n$, $j = 1, 2, \dots, m$ form a basis of L over F and the result follows. ■

Definition 13.13. Let S be a subset of a field K , then a subfield K' of K is said to be generated by S if

(i) $S \subseteq K'$.

(ii) For any subfield L of K , $S \subseteq L$ implies $K' \subseteq L$.

NOTATION. Subfield generated by a subset S will be denoted by $\langle S \rangle$.

Essentially the subfield generated by S is the intersection of all subfields of K which contain S . Now let K be a field extension of F and S be any subset of K , then the subfield of K generated by $F \cup S$ is said to be the subfield of K generated by S over F and this subfield is denoted by $F(S)$. However if S is a finite set and its members are $a_1, a_2, \dots, a_n$ then we write $F(S) = F(a_1, a_2, \dots, a_n)$.

Definition 13.14. A field K is said to be *finitely generated over F* if there exists a finite number of elements $a_1, a_2, \dots, a_n$ in K such that $K = F(a_1, a_2, \dots, a_n)$.

In particular if K is generated by a single element over F , then K is called a *simple extension* of F . Thus $\mathbb{Q}[\sqrt{2}] = \mathbb{Q}(\sqrt{2})$ is a simple extension of $\mathbb{Q}$. Consider any $a \in K$ and let $F(a)$ be the subfield of K generated by a over F . Then for any $\alpha_0, \alpha_1, \alpha_2, \dots, \alpha_n \in F$, $\alpha_0 + \alpha_1 a + \alpha_2 a^2 + \dots + \alpha_n a^n \in F(a)$. This means that $F[a] \subseteq F(a)$. Consequently the field of quotients T of $F[a]$ is also contained in $F(a)$. However $F \subseteq T$ and $a \in T$ therefore $F(a) \subseteq T$ and we get $T = F(a)$ i.e., $F(a)$ is the field of quotients of the subring of K generated by $F \cup \{a\}$. This discussion can be extended to any set S and we can say $F(S)$ is the field of quotients of $F[S]$, the subring of K generated by $F \cup S$. Let us remark that for any $f(x) = \alpha_0 + \alpha_1 x + \dots + \alpha_n x^n$ in $F[x]$ and for any $a \in K$, $f(a)$ denotes $\alpha_0 + \alpha_1 a + \dots + \alpha_n a^n$. It can also be verified that the mapping $f(x) \rightarrow f(a)$ is a ring homomorphism of $F[x]$ into K .

Definition 13.15. An element a of K is said to be a *root* or *zero* of a polynomial $p(x) \in F[x]$ if $p(a) = 0$. In this case we also say that $p(x)$ is *satisfied by a* .

Definition 13.16. An element a of K is said to be *algebraic over F* if a is a root of a non-zero polynomial $f(x) \in F[x]$. K is said to be an *algebraic extension* of F if every element of K is algebraic over F .

If $f(x) = \alpha_0 + \alpha_1 x + \dots + \alpha_n x^n$ with $\alpha_0 \neq 0$, then $f(a) = \alpha_0 + \alpha_1 a + \dots + \alpha_n a^n$. So, if $f(a) = 0$ we cannot have $n = 0$. Thus $n > 0$ and $\deg f(x) > 0$ i.e. $f(x)$ is a non-constant polynomial.

Definition 13.17. A non-zero polynomial $f(x)$ in $F[x]$ is said to be a *monic polynomial over F* if the coefficient of highest power of x in $f(x)$ is equal to 1, the unity of F .

Theorem 13.18. If an element $a \in K$ is algebraic over F , then there exists a unique monic polynomial $p(x)$ of positive degree over F , such that (i) $p(a) = 0$ (ii) if for any $f(x) \in F[x]$, $f(a) = 0$ then $p(x)$

divides $f(x)$.

Proof: Since a is algebraic over K , a is a root of some non-zero polynomial over F . Let $t(x)$ be a non-zero polynomial over F of smallest degree such that $t(a)=0$. Let degree of $t(x)$ be equal to n . Then $n > 0$ and $t(x) = \alpha_0 + \alpha_1 x + \alpha_2 x^2 + \dots + \alpha_n x^n$ with $\alpha_i \in F$ and $\alpha_n \neq 0$. We take

$p(x) = x^n + \alpha_{n-1} x^{n-1} + \alpha_{n-2} x^{n-2} + \dots + \alpha_0 = \alpha_n^{-1} t(x)$ and find that $p(a)=0$, with $\deg p(x) = n = \deg t(x)$. This shows that $p(x)$ is a monic polynomial over F such that $p(x)$ has smallest degree among all those non-zero polynomials over F which are satisfied by a .

Let $f(x)$ be any polynomial such that $f(a)=0$. By division algorithm, we can find polynomials $q(x), r(x)$ in $F[x]$, such that

$$f(x) = p(x)q(x) + r(x)$$

and

$$r(x) = 0 \text{ or } \deg r(x) < \deg p(x).$$

Now

$$0 = f(a) = p(a)q(a) + r(a) = r(a).$$

Consequently $r(x) = 0$, as no non-zero polynomial of degree less than $\deg p(x)$ is satisfied by a . So $f(x) = p(x)q(x)$ and $p(x)$ divides $f(x)$. Suppose now that $p_1(x) = x^m + \beta_{m-1}x^{m-1} + \dots + \beta_0$ be another monic polynomial such that $p_1(a)=0$ and for any polynomial $f(x)$ in $F[x]$, $f(a)=0$ implies $p_1(x)$ divides $f(x)$. Then in particular $p_1(x) \mid p(x)$ so $p(x) = l(x)p_1(x)$ for some $l(x) \in F[x]$. Then the minimality of degree of $p(x)$ gives that $l(x)$ is a non-zero constant. Let $l(x) = c$ and $\deg p(x) = n$. Then comparing the coefficient of x^n in $p(x)$ and $cp_1(x)$ we get $c=1$. That gives $p(x) = p_1(x)$. Hence $p(x)$ is unique. This completes the theorem. ■

Definition 13.19. If any element $a \in K$ is algebraic over F , then a monic polynomial of smallest degree over F satisfied by a is called the *minimal polynomial* of a over F .

If the degree of the minimal polynomial of a is n then a is said to be algebraic over F of degree n .

EXAMPLE 5. We assert that $x^2 - 3$ is the minimal polynomial of $\sqrt{3}$ over $\mathbb{Q}$. Clearly $x^2 - 3$ is a monic polynomial and is satisfied by $\sqrt{3}$. As $\sqrt{3}$ is irrational it can't satisfy a polynomial of degree 1 over $\mathbb{Q}$. So our assertion follows.

EXAMPLE 6. $x^3 - 2$ is the minimal polynomial of $\sqrt[3]{2}$ over $\mathbb{Q}$. $x^3 - 2$ is a monic polynomial satisfied by $\sqrt[3]{2}$. By Eisenstein's Criterion $x^3 - 2$ is irreducible over $\mathbb{Q}$ so no polynomial over $\mathbb{Q}$ of degree less than 3 can be satisfied by $\sqrt[3]{2}$ (Theorem 13.18). Thus $x^3 - 2$ is the minimal polynomial of $\sqrt[3]{2}$ over $\mathbb{Q}$.

Theorem 13.18 shows that the minimal polynomial of $a \in K$ over F is unique.

Recall the definition of irreducible element over an integral

domain D with unity. We claim that the minimal polynomial $p(x)$ of an algebraic element a of K over F is an irreducible element. Firstly notice that $\deg p(x) > 0$, so $p(x)$ is not a unit. Let $p(x) = q(x)s(x)$ for some non-zero polynomials $q(x), s(x)$ in $F[x]$. Then $0 = p(a) = q(a)s(a)$ gives $q(a) = 0$ or $s(a) = 0$. So $p(x) \mid q(x)$ or $p(x) \mid s(x)$. Since $q(x) \mid p(x)$ and $s(x) \mid p(x)$, $p(x)$ is either associate of $q(x)$ or an associate of $s(x)$. This shows that $p(x)$ is an irreducible element. Further recall that in a domain D which is a PID if an element p of D is irreducible then for any $a \in D$ either $p \mid a$ or the HCF of a and p is 1 i.e. $p\lambda + a\mu = 1$ for some $\lambda, \mu \in D$. In particular for any $f(x)$ in $F[x]$ either $p(x) \mid f(x)$ or there exist two polynomials $A(x), B(x)$ in $F(x)$ such that $p(x)A(x) + f(x)B(x) = 1$.

We now prove the following theorem which is of far reaching consequence.

Theorem 13.20. *An element a of K is algebraic over F if and only if $[F(a) : F]$ is finite.*

Proof: Let $[F(a) : F] = n$, a positive integer. Then any $n+1$ elements of $F(a)$ are linearly dependent over F . In particular $1, a, a^2, \dots, a^n$ are linearly dependent over F . So there exist $\alpha_0, \alpha_1, \dots, \alpha_n \in F$ not all zero, such that

$$\alpha_0 + \alpha_1 a + \alpha_2 a^2 + \dots + \alpha_n a^n = 0.$$

This proves that a is algebraic over F .

Conversely let a be algebraic over F and $p(x)$ be the minimal polynomial of a over F . Let $n = \deg p(x)$. Consider any non-zero element $f(a) = \alpha_0 + \alpha_1 a + \dots + \alpha_m a^m$ in $F[a]$. Then $p(x)$ does not divide $f(x)$. Consequently there exist $A(x)$ and $B(x)$ such that

$$p(x)A(x) + f(x)B(x) = 1 \quad (\text{See the remarks preceding this theorem}).$$

This gives $1 = p(a)A(a) + f(a)B(a) = f(a)E(a)$ as $p(a) = 0$.

So $[f(a)]^{-1} = B(a) \in F[a]$. Hence $F[a]$ is a field.

However $F(a)$ is the field of quotients of $F[a]$. We get $F(a) = F[a]$. Now consider the elements $1, a, a^2, \dots, a^{n-1}$ in $F(a)$. If for some elements $\alpha_0, \alpha_1, \alpha_2, \dots, \alpha_{n-1} \in F$ not all zero,

$$\alpha_0 1 + \alpha_1 a + \alpha_2 a^2 + \dots + \alpha_{n-1} a^{n-1} = 0$$

then a satisfies a non-zero polynomial of degree $< n = \text{degree of the minimal polynomial}$. This is a contradiction. Hence $1, a, a^2, \dots, a^{n-1}$ are linearly independent over F .

Consider any $f(a) = \beta_0 + \beta_1 a + \beta_2 a^2 + \dots + \beta_m a^m \in F[a] = F(a)$ where $f(x) = \beta_0 + \beta_1 x + \dots + \beta_m x^m$. By the division algorithm there exist polynomials $q(x)$ and $r(x)$ such that

$$f(x) = q(x)p(x) + r(x) \text{ with either } r(x) = 0 \text{ or } \deg r(x) < \deg p(x).$$

Now $f(a) = q(a)p(a) + r(a) = r(a)$, as $p(a) = 0$.

Since either $r(x) = 0$ or $\deg r(x) < \deg p(x)$, we have

$r(x) = \delta_0 + \delta_1 x + \dots + \delta_{n-1} x^{n-1}$ for some $\delta_i \in F$.

Thus $f(a) = r(a) = \delta_0 + \delta_1 a + \dots + \delta_{n-1} a^{n-1}$. This proves that $f(a)$ is a linear combination of $1, a, \dots, a^{n-1}$. Hence $\{1, a, a^2, \dots, a^{n-1}\}$ is a basis of $F(a)$ over F . Consequently $[F(a) : F] = n$, which is finite. ■

Theorem 13.21. Every finite extension of a field is algebraic extension.

Proof: Let K be a finite extension of a field F . Let $[K : F] = n$. For any $a \in K$, $F(a)$ is a subfield of K .

But $[F(a) : F] \mid n$ so it is finite. Hence a is algebraic over F (Theorem 13.20). Consequently K is algebraic extension of F . ■

REMARK 1. Converse of the above theorem is not true, in this connection see Example 15.

REMARK 2. Suppose F_1 be a field such that $F \subseteq F_1 \subseteq K$ and an element a in K be algebraic over F . Since every polynomial over F is a polynomial over F_1 , it follows that a is algebraic over F_1 also. So if $p(x)$ is the minimal polynomial of a over F and $p_1(x)$ is the minimal polynomial of a over F_1 then as $p(x) \in F_1[x]$ we get $p_1(x) \mid p(x)$ in $F_1[x]$. That gives $\deg p_1(x) \leq \deg p(x)$. However we know that

$$[F(a) : F] = \deg p(x) \text{ and } [F_1(a) : F_1] = \deg p_1(x).$$

So we have $[F_1(a) : F_1] \leq [F(a) : F]$.

Now we establish an important theorem.

Theorem 13.22. If L is an algebraic extension of K and K is an algebraic extension of F , then L is an algebraic extension of F .

Proof: Let $a \in L$, since L is an algebraic extension of K

$a^n + \alpha_1 a^{n-1} + \alpha_2 a^{n-2} + \dots + \alpha_n = 0$ holds for some $\alpha_i \in K (i=1, 2, \dots, n)$ and $n \geq 1$.

Define $F_0 = F$ and $F_i = F_{i-1}(\alpha_i)$ for $1 \leq i \leq n$.

Now each α_i is algebraic over F , so α_i is also algebraic over F_{i-1} (since $F \subseteq F_{i-1} \subseteq K$) and $[F_i : F_{i-1}] \leq [F(\alpha_i) : F]$ (By the Remark 2 above).

This gives $[F_i : F_{i-1}]$ is finite.

The relation $[F_n : F] = [F_n : F_{n-1}][F_{n-1} : F_{n-2}] \dots [F_1 : F]$ gives that $[F_n : F]$ is finite. Now $\alpha_i \in F_n \forall i=1, 2, \dots, n$. Also a is algebraic over F_n as it satisfies a non-zero polynomial with coefficients in F_n . Hence $[F_n(a) : F_n]$ is finite.

Then $[F_n(a) : F] = [F_n(a) : F_n][F_n : F]$ implies that $[F_n(a) : F]$ is finite. So $F_n(a)$ is an algebraic extension of F (Theorem 13.21). Consequently a is algebraic over F as $a \in F_n(a)$.

Hence L is an algebraic extension of F . ■

Theorem 13.23. The set S of all those elements of K which are algebraic over F is a subfield of K containing F such that no element a of K which is not in S is algebraic over S .

Proof: Since each $\alpha \in F$ satisfies the polynomial $x - \alpha$ over F , and so $\alpha \in S$ hence $F \subseteq S$. Consider $a, b \in S$ then $[F(a) : F]$ and $[F(b) : F]$ are both finite. As b is algebraic over F , b is *a fortiori* algebraic over $F(a)$; consequently $[F(a, b) : F(a)]$ is finite. This implies $[F(a, b) : F] = [F(a, b) : F(a)][F(a) : F]$ is finite. Hence $F(a, b)$ is an algebraic extension of F . Now $a - b \in F(a, b)$, so $a - b$ is algebraic over F and $a - b \in S$. Similarly if $b \neq 0$ then $ab^{-1} \in S$. Hence S is a subfield of K containing F . This completes first part of the theorem. If in the above discussion one replaces F by S then S_1 , the set of all those elements of K which are algebraic over S form a subfield of K containing S . Thus S_1 is an algebraic extension of S and S is an algebraic extension of F . So S_1 is also an algebraic extension of F (Theorem 13.22). Consequently $S_1 \subseteq S$ and hence $S = S_1$. If any $a \in K$ is algebraic over S then by definition of S_1 , $a \in S_1$. So $a \in S$ as $S_1 = S$. Hence no element of K not in S is algebraic over S . ■

Corollary 13.23(a). If any two elements a, b of K are algebraic over F then

- (i) $a \pm b, ab$ are algebraic over F .
- (ii) If $b \neq 0$ then ab^{-1} is algebraic over F .

Proof follows from the fact that $a - b, ab^{-1}$ are in S and S is an algebraic extension of F . ■

One can interpret above theorem by saying that S is the largest possible algebraic extension of F within K , and S is algebraically closed with respect to K in the sense that S has no proper algebraic extension in K . This motivates us to call S as the algebraic closure of F relative to K . In particular when K itself is an algebraic extension of F then $S = K$.

Let us have a closer look at the field $F(a)$, when $a \in K$ is algebraic over F . It was shown earlier (Theorem 13.18) that if $p(x)$ is the minimal polynomial of a over F then for any $f(x) \in F[x]$, $f(a) = 0$ if and only if $f(x) \in \langle p(x) \rangle$. Now the mapping $\sigma : F[x] \rightarrow F[a]$ defined by $\sigma[g(x)] = g(a)$ for all $g(x) \in F[x]$ is an onto homomorphism and clearly $g(x) \in \text{Ker } \sigma$ if and only if $g(a) = 0$. Thus $\text{Ker } \sigma = \langle p(x) \rangle$. Hence by the Fundamental Theorem of Homomorphism, we get

$\frac{F[x]}{\langle p(x) \rangle} \cong F[a]$. However $F[a] = F(a)$ (as proved during the proof

of Theorem 13.20). Hence $\frac{F[x]}{\langle p(x) \rangle} \cong F(a)$. Thus we get

Theorem 13.24. If any element a of K is algebraic over F and if $p(x)$ is the minimal polynomial of a over F then $\frac{F(x)}{\langle p(x) \rangle} \cong F(a)$. ■

REMARK. In the isomorphism $\eta : \frac{F[x]}{\langle p(x) \rangle} \rightarrow F(a)$ we see that $\eta[\overline{g(x)}] = g(a) \vee g(x) \in F[x]$. Thus in particular for any $\alpha \in F$, $\eta(\overline{\alpha}) = \alpha$ [Since if $g(x) = \alpha$ then $g(a) = \alpha$].

Theorem 13.25. *If any two elements a and b of K satisfy same minimal polynomial $p(x)$ over F then there exists an isomorphism σ of $F(a)$ onto $F(b)$ such that $\sigma(\alpha) = \alpha$ for all $\alpha \in F$ and $\sigma(a) = b$, i.e., σ leaves F elementwise fixed and maps a onto b .*

Proof: By Theorem 13.24 there exist isomorphism η_1 from $\frac{F[x]}{\langle p(x) \rangle}$ onto $F(a)$ and η_2 from $\frac{F[x]}{\langle p(x) \rangle}$ onto $F(b)$ such that for any $f(x) \in F[x]$, $\eta_1[\overline{f(x)}] = f(a)$ and $\eta_2[\overline{f(x)}] = f(b)$. Then $\sigma = \eta_2 \eta_1^{-1}$ is an isomorphism of $F(a)$ onto $F(b)$ such that

$$\sigma(\alpha) = \eta_2 \eta_1^{-1}(\alpha) = \eta_2(\overline{\alpha}) = \alpha \quad \forall \alpha \in F$$

$$\text{and} \quad \sigma(a) = \eta_2 \eta_1^{-1}(a) = \eta_2(x) = b. \blacksquare$$

Lemma 13.26. Factor Theorem.

Any element $a \in K$ is root of a polynomial $f(x)$ over F of positive degree if and only if $(x-a) \mid f(x)$ in $K[x]$.

Proof: Since $f(x) \in K[x]$, $x-a \in K[x]$, by division algorithm there exist unique polynomials $q(x)$, $r(x) \in K[x]$ such that $f(x) = (x-a)q(x) + r(x)$ and either $r(x) = 0$ or $\deg r(x) < \deg(x-a) = 1$. Consequently $r(x) = c$ for some $c \in K$ and we get $f(x) = (x-a)q(x) + c$. . . (1)

Let a be a root of $f(x)$ then $f(a) = 0$. Substituting a for x in (1) we get $0 = f(a) = (a-a)q(a) + c$ i.e., $c = 0$.

So $f(x) = (x-a)q(x)$ and $(x-a) \mid f(x)$.

Conversely if $(x-a) \mid f(x)$ in $K[x]$ then $f(x) = (x-a)g(x)$ for some $g(x) \in K[x]$. Therefore $f(a) = (a-a)g(a) = 0$ and a is a root of $f(x)$. $\blacksquare$

Definition 13.27. If an element a of a field K is a root of a non-zero polynomial $f(x)$ over a subfield F of K , then a is said to be a root of multiplicity m in K if $(x-a)^m \mid f(x)$ but $(x-a)^{m+1} \nmid f(x)$ in $K[x]$. $\blacksquare$

Thus if a is a root of multiplicity m in K then $f(x) = (x-a)^m q(x)$ for some $q(x) \in K[x]$. If a is a root of $q(x)$ then by Lemma 13.26, $(x-a) \mid q(x)$ and this in turn implies that $(x-a)^{m+1} \mid f(x)$. This contradicts the assumption that a is a root of multiplicity m in K . Hence $q(a) \neq 0$.

Following questions about roots are extremely pertinent.

1. How many roots a polynomial $f(x)$ of positive degree in $F[x]$ can have in any field extension of F ?

2. Does there exist a field extension of F in which $f(x)$ has largest

possible number of roots and what is this largest number?

Question 1 is answered by the following theorem :

Theorem 13.28. *A polynomial of degree $n \geq 1$ over a field F cannot have more than n roots in any field extension of F .*

Proof: Let $f(x)$ be a polynomial of degree $n \geq 1$ in $F[x]$. We apply induction on n . Let $n=1$. Thus $f(x)=\alpha x + \beta$ for some $\alpha, \beta \in F$ and $\alpha \neq 0$. Then $-\beta\alpha^{-1}$ is the only root of $f(x)$ in any field extension of F . So the result holds for $n=1$. To apply induction, we suppose that $n > 1$ and the result holds for all polynomials of positive degree $< n$ over any field. Let K be a field extension of F . If K has no roots of $f(x)$, the result holds trivially. Suppose some $a \in K$ is a root of $f(x)$, of multiplicity $m \geq 1$ in K . Then $f(x) = (x-a)^m f_1(x)$ for some $f_1(x) \in K[x]$. Now $\deg f_1(x) = n-m < n$. So $f_1(x)$ cannot have more than $n-m$ roots in K . Consequently $f(x)$ cannot have more than $n = (n-m) + m$ roots in K . ■

After having answered first question, we turn our attention to the next question. Let us recall that if any $a \in K$ is algebraic over F then the minimal polynomial $p(x)$ of a over F is irreducible and further $\deg p(x) \geq 2$ whenever $a \notin F$. Firstly we take the converse of the above result in the following.

Theorem 13.29. (Kronecker). *If $p(x)$ is any irreducible polynomial over F then there exists an extension E of F such that $[E:F] = \deg p(x)$ and $p(x)$ has a root in E .*

Proof: Since $p(x)$ is irreducible, $V = \langle p(x) \rangle$ is a maximal ideal of $F[x]$. Thus $\frac{F[x]}{V}$ is a field. Let us take $E = \frac{F[x]}{V}$.

Define a mapping $\sigma : F \rightarrow E$ by $\sigma(\alpha) = \bar{\alpha} = \alpha + V \in E \forall \alpha \in F$.

σ is a ring homomorphism. Since any non-zero member of $V = \langle p(x) \rangle$ is of degree ≥ 1 , $\bar{\alpha} \neq \bar{0}$ whenever $\alpha \neq 0$.

Hence σ is a monomorphism. Consequently (E, σ) is an extension of F . As per our convention we identify F with its image $\sigma(F)$ in E and each α in F with $\bar{\alpha}$ in E . Let $p(x) = \alpha_0 + \alpha_1 x + \dots + \alpha_n x^n$, $\alpha_n \neq 0$, $\alpha_i \in F$. Then $\bar{0} = \overline{p(x)} = \alpha_0 + \alpha_1 \bar{x} + \alpha_2 (\bar{x})^2 + \dots + \alpha_n (\bar{x})^n$ gives $\bar{x}$ is a root of $p(x)$. Thus E contains a root of $p(x)$. This gives immediately that $[E:F] \geq \deg p(x) = n$.

Now for any $f(x) \in F[x]$ we can write $f(x) = p(x)q(x) + r(x)$ for some $q(x)$, $r(x) \in F[x]$ with $r(x) = 0$ or $\deg r(x) < \deg p(x)$. We get $\overline{f(x)} = \overline{r(x)}$ as $\overline{p(x)} = \bar{0}$.

As $r(x) = 0$ or $\deg r(x) < \deg p(x)$ we can write

$r(x) = \beta_0 + \beta_1 x + \dots + \beta_{n-1} x^{n-1}$ for some $\beta_i \in F$, $i = 0, 1, 2, \dots, n-1$.

So $\overline{f(x)} = \overline{r(x)} = \beta_0 + \beta_1 \bar{x} + \dots + \beta_{n-1} (\bar{x})^{n-1}$. This shows that every member of E is a linear combination of $1, \bar{x}, (\bar{x})^2, \dots, (\bar{x})^{n-1}$.

So $[E : F] \leq n$. As proved earlier $[E : F] \geq n$, we get $[E : F] = n$.
Now we are ready to give complete answer to Question 2.

Theorem 13.30. *If $f(x)$ is any polynomial of degree $n \geq 1$ over a field F then there exists a field extension E of F such that $f(x)$ has n roots in E and $[E : F] \leq n!$.*

Proof: The proof is based upon induction on degree of the polynomial $f(x)$. Let $\deg f(x) = 1$ then $f(x)$ is of the form $\alpha x + \beta$ for some $\alpha, \beta \in F$ with $\alpha \neq 0$ and its only root is $-\beta\alpha^{-1}$ which belongs to F . Here we can take $E = F$ and thus we see that the result holds in case $\deg f(x) = 1$. To apply induction, suppose that the result holds for any polynomial of degree $< n$ over any field. Let $p(x)$ be an irreducible factor of $f(x)$ and let its degree be m . Then $m \leq n$ and there exists a field extension E' of F such that $p(x)$ has a root, say, a in E' and $[E' : F] = m$ (Theorem 13.29). Then a is also a root of $f(x)$ and $f(x) = (x - a)f_1(x)$ for some $f_1(x) \in E'[x]$ (Why?). Now $\deg f_1(x) = n - 1 < n$. Hence by induction hypothesis there exists a field extension E of E' such that $f_1(x)$ has $n - 1$ roots in E and $[E : E'] \leq (n - 1)!$. Also $a \in E$. Since the n roots of $f(x)$ are a and the $n - 1$ roots of $f_1(x)$ we get $f(x)$ has all its n roots in E . Further $[E : F] = [E : E'] [E' : F] \leq (n - 1)! n = n!$.

REMARK. In the above theorem if $a_1, a_2, \dots, a_n$ are the n roots of $f(x)$ in K then $F(a_1, a_2, \dots, a_n)$ is the subfield of E containing F such that there does not exist any proper subfield K of $F(a_1, a_2, \dots, a_n)$ containing F and all roots of $f(x)$. Further a simple inductive argument establishes that $f(x) = \alpha(x - a_1)(x - a_2) \dots (x - a_n)$ for some $\alpha (\neq 0) \in F$. Thus $f(x) = (\alpha x - \alpha a_1)(x - a_2) \dots (x - a_n)$ is a product of n linear factors.

Definition 13.31. Let $f(x)$ be any polynomial of degree $n \geq 1$ over a field F . Then a field extension E of F is called a *splitting field* of $f(x)$ if

- $f(x)$ can be factored into n linear factors over E , and
- there does not exist any proper subfield E' of E containing F such that $f(x)$ can be factored into n linear factors over E' .

Equivalently one can say that E is a splitting field of $f(x)$, if E contains roots of $f(x)$ and $E = F(a_1, a_2, \dots, a_n)$, the field generated by F and n roots $a_1, a_2, \dots, a_n$, of $f(x)$ in E . The above theorem shows the existence of splitting field. This theorem does not say anything whether there is any relation between two splitting fields of the same polynomial or not. This we take up now.

Recall that if R and R' are any two rings such that there exists an isomorphism σ of R onto R' then σ can be extended to an isomorphism τ of $R[x]$ onto $R'[t]$ where x and t are indeterminates such that

$\tau(x)=t$. If A is any ideal of $R[x]$ then $\tau(A)$ is an ideal of $R'[t]$, and we can define an isomorphism $\eta: \frac{R[x]}{A} \rightarrow \frac{R'[t]}{\tau(A)}$

by $\eta[f(x)+A]=\tau[f(x)]+\tau(A)$ for all $f(x) \in R[x]$. Now for any $f(x) = \alpha_0 + \alpha_1 x + \alpha_2 x^2 + \dots + \alpha_n x^n \in R[x]$, $\tau[f(x)] = \sigma(\alpha_0) + \sigma(\alpha_1)t + \dots + \sigma(\alpha_n)t^n$.

In particular we take two fields F and F' and take σ as an isomorphism of F onto F' and τ the corresponding extension which is an isomorphism of $F[x]$ onto $F'[t]$. For our convenience, for each $\alpha \in F$ we denote its image $\sigma(\alpha)$ by α' and for any $f(x) \in F[x]$ we denote by $f'(t)$ the image $\tau[f(x)]$.

Now let $p(x)$ be any irreducible polynomial over F . Suppose $p'(t) = r'(t)s'(t)$ for some polynomials $r'(t), s'(t)$ of positive degree in $F'[t]$. Since τ is an isomorphism we get $r(x), s(x) \in F[x]$ such that $r'(t)$ and $s'(t)$ are respective images of $r(x)$ and $s(x)$ under τ . Notice that τ preserves degrees, so $r(x)$ and $s(x)$ polynomials of positive degree such that

$$\tau[p(x)] = \tau[r(x)]\tau[s(x)] = \tau[r(x)s(x)]$$

Consequently $p(x) = r(x)s(x)$ and $p(x)$ is reducible.

This contradicts the assumption that $p(x)$ is irreducible. Hence $p'(t)$ is also irreducible. As explained earlier there exists an isomorphism

$$\eta: \frac{F[x]}{\langle p(x) \rangle} \rightarrow \frac{F'[t]}{\langle p'(t) \rangle}$$

in which $\eta[f(x) + \langle p(x) \rangle] = f'(t) + \langle p'(t) \rangle$ for all $f(x) \in F[x]$. Now we prove the following.

Theorem 13.32. Let $p(x)$ be an irreducible polynomial in $F[x]$ and $p'(t)$, the corresponding polynomial in $F'[t]$. Suppose u and v are roots of $p(x)$ and $p'(t)$ respectively in some field extensions E and E' of F and F' respectively, then there exists an isomorphism μ of $F(u)$ onto $F'(v)$ such that $\mu(\alpha) = \alpha' \forall \alpha \in F$ and $\mu(u) = v$.

Proof: Now $p'(t)$ is also irreducible (by remarks preceding this theorem). There exist isomorphisms

$$\sigma_1: \frac{F[x]}{\langle p(x) \rangle} \rightarrow F(u)$$

$$\text{and } \sigma_2: \frac{F'[t]}{\langle p'(t) \rangle} \rightarrow F'(v)$$

such that $\sigma_1[f(x) + \langle p(x) \rangle] = f(u)$ and $\sigma_2[f'(t) + \langle p'(t) \rangle] = f'(v)$, (Theorem 13.24 and the remark).

$$\text{Further } \eta: \frac{F[x]}{\langle p(x) \rangle} \rightarrow \frac{F'[t]}{\langle p'(t) \rangle}$$

defined by $\eta[f(x) + \langle p(x) \rangle] = f'(t) + \langle p'(t) \rangle$ is an isomorphism. Notice that $\sigma_1[x + \langle p(x) \rangle] = u$ and $\sigma_2[t + \langle p'(t) \rangle] = v$.

Now $\mu = \sigma_2 \eta \sigma_1^{-1}$ is an isomorphism of $F(u)$ onto $F'(v)$.

$$\begin{aligned}\text{For any } \alpha \in F, \mu(\alpha) &= (\sigma_2 \eta \sigma_1^{-1})(\alpha) \\ &= \sigma_2 \eta[\alpha + \langle p(x) \rangle] \\ &= \sigma_2[\alpha' + \langle p'(t) \rangle] \\ &= \alpha'.\end{aligned}$$

$$\begin{aligned}\text{and } \mu(u) &= (\sigma_2 \eta \sigma_1^{-1})(u) = \sigma_2 \eta[x + \langle p(x) \rangle] \\ &= \sigma_2[t + \langle p'(t) \rangle] = v. \blacksquare\end{aligned}$$

Corollary 13.33. Let $F \subseteq K$ be any two fields. If $\alpha, \beta \in K$ have same minimal polynomial $p(x)$ over F , then there exists an isomorphism μ of $F(\alpha)$ onto $F(\beta)$ such that $\mu(\alpha) = \beta$ and $\mu(a) = a$ for every $a \in F$.

Proof: In the above theorem take $F = F'$ and σ , the identity mapping. The result follows. ■

Definition. 13.34. Two elements a, b of field K are said to be conjugate over a subfield F of K if they are algebraic over F and have the same minimal polynomial over F .

EXAMPLE 7. $\sqrt{3}$ and $-\sqrt{3}$ are conjugate elements of $\mathbf{R}$ over $\mathbf{Q}$ as $x^2 - 3$ is the minimal polynomial over $\mathbf{Q}$ of both $\sqrt{3}$ and $-\sqrt{3}$.

EXAMPLE 8. $\sqrt[3]{2}, \sqrt[3]{2}\omega, \sqrt[3]{2}\omega^2$ where $\omega^2 + \omega + 1 = 0$, are elements of $\mathbf{C}$ conjugate over $\mathbf{Q}$. All of them have the same minimal polynomial $x^3 - 2$ over $\mathbf{Q}$.

Theorem 13.35. Let $F \cong F'$ and let $f(x)$ be any polynomial of degree ≥ 1 over F and $f'(t)$ be the corresponding polynomial over F' . If E and E' are splitting fields of $f(x)$ and $f'(t)$ over F and F' respectively, then there exists an isomorphism ψ of E onto E' such that $\psi(\alpha) = \alpha' \forall \alpha \in F$.

Proof: Let $\deg f(x) = n$. Notice that the degree $[E : F]$ is finite. We apply induction on $[E : F]$. Suppose $[E : F] = 1$ then $E = F$ and all the n roots $a_1, a_2, \dots, a_n$ of $f(x)$ are in F , further

$$f(x) = \alpha(x - a_1)(x - a_2) \dots (x - a_n)$$

for some $\alpha (\neq 0) \in F$. So $f'(t) = \alpha'(t - a'_1)(t - a'_2) \dots (t - a'_n)$

gives $a'_1, a'_2, \dots, a'_n$ are all in F' . Hence $E' = F'$. We can take ψ to be just the isomorphism σ of F onto F' . So the result holds when $[E : F] = 1$. To apply induction let $[E : F] > 1$ and let us suppose that the result holds for all splitting fields E over any field F of degree $< [E : F]$. Since $[E : F] > 1$, at least one root, say u , of $f(x)$ is not in F . If $p(x)$ is the minimal polynomial of u over F , then $\deg p(x) > 1$. Let $p'(t)$ be the corresponding polynomial over F' . $p'(t)$ is

irreducible. As $p(x) \mid f(x)$ we get $f(x) = p(x)q(x)$ for some $q(x) \in F[x]$ and therefore $f'(t) = p'(t)q'(t)$. Let v be a root of $p'(t)$ in E' . There exists an isomorphism

$$\theta : F(u) \rightarrow F'(v)$$

in which $\theta(\alpha) = \sigma(\alpha) = \alpha' \forall \alpha \in F$ and $\theta(u) = v$ (Theorem 13.32).

Now $f(x) = (x-u)f_1(x)$ and $f'(t) = (t-v)g_1(t)$ for some $f_1(x) \in F(u)[x]$ and $g_1(t) \in F'(v)[t]$. If θ_1 denotes the isomorphism of $F(u)[x]$ onto $F'(v)[t]$ such that θ_1 is an extension of θ and $\theta_1(x) = t$ then $f'(t) = \theta_1[f(x)] = \theta_1[(x-u)f_1(x)] = (t-v)\theta_1[f_1(x)]$ gives $\theta_1[f_1(x)] = g_1(t)$ as $f'(t) = (t-v)g_1(t)$.

Further as E contains all the n roots of $f(x)$, E contains all the $n-1$ roots of $f_1(x)$. If E_1 is a splitting field of $f_1(x)$ over $F(u)$ contained in E , then, besides having all roots of $f_1(x)$, E_1 contains u also. Consequently E_1 contains all roots of $f(x)$. So by definition of splitting field we get $E \subseteq E_1$. Hence $E = E_1$. Thus E is splitting field of $f_1(x)$ over $F(u)$. Similarly E' is a field of $g_1(t)$ over $F'(v)$.

Now $[E : F] = [E : F(u)][F(u) : F]$ and $[F(u) : F] = \deg p(x) > 1$.

That gives $[E : F(u)] < [E : F]$. Hence by induction hypothesis there exists an isomorphism ψ of E onto E' which extends the isomorphism $\theta : F(u) \rightarrow F'(v)$. However θ extends σ . Hence we see that $\psi(\alpha) = \theta(\alpha) = \alpha' \forall \alpha \in F$. This proves the theorem. ■

In particular if we take $F = F'$ and σ to be the identity map, we conclude the following:

Corollary 13.36. *If F and E' are any two splitting fields of same polynomial $f(x)$ of degree $n \geq 1$ over a field F , then there exists an isomorphism ψ of E onto E' such that $\psi(\alpha) = \alpha$ for all $\alpha \in F$.*

Let a be an element of K algebraic over F and $p(x)$ be the minimal polynomial of a over F . For any non zero $f(a) \in F(a)$, where $f(x)$ is a polynomial over F , we have shown (Theorem 13.20) that the inverse of $f(a)$ is some polynomial $g(a) \in F(a)$. We illustrate this here.

EXAMPLE 9. Consider the field $\mathbb{Q}(\sqrt{2})$.

$$3 + \sqrt{2} \in \mathbb{Q}(\sqrt{2}). \text{ Then } \frac{1}{3 + \sqrt{2}} = \frac{3 - \sqrt{2}}{7} = \frac{3}{7} - \frac{1}{7}\sqrt{2}$$

is a polynomial in $\sqrt{2}$ over $\mathbb{Q}$.

EXAMPLE 10. Consider $1 - \sqrt[3]{2} + \sqrt[3]{4} \in \mathbb{Q}[\sqrt[3]{2}] = \mathbb{Q}(\sqrt[3]{2})$. Let $f(x) = x^3 - x + 1$, $p(x) = x^3 - 2$, $p(x)$ is the minimal polynomial of $\sqrt[3]{2}$. To find the inverse of $f(\sqrt[3]{2})$ we find two polynomials $g(x)$ and $h(x)$ such that $f(x)g(x) + p(x)h(x) = 1$. These $g(x)$ and $h(x)$ are found by the method of division algorithm. Now $\deg f(x) < \deg p(x)$, dividing $p(x)$ by $f(x)$ we obtain $x^3 - 2 = (x+1)(x^2 - x + 1) - 3$.

Thus $(x^3-2)(-1/3)+(x+1)(x^2/3-x/3+1/3)=1$
 Hence replacing x by $\sqrt[3]{2}$ we get

$$\frac{(\sqrt[3]{2}+1)}{3}(\sqrt[3]{4}-\sqrt[3]{2}+1)=1$$

So that $(\sqrt[3]{4}-\sqrt[3]{2}+1)^{-1}=\frac{1}{3}+\frac{\sqrt[3]{2}}{3}$, a polynomial in $\sqrt[3]{2}$ over $\mathbb{Q}$.

Now we give examples of two distinct monic irreducible polynomials having same splitting field.

EXAMPLE 11. Consider x^2+3 , x^2+x+1 over $\mathbb{Q}$. These quadratics have no rational roots, so they are irreducible (Note in general if $f(x)$ is a polynomial of degree >3 over a field F such that $f(x)$ has no root in F we cannot say that $f(x)$ is irreducible). The roots of x^2+3 are $\pm\sqrt{-3}$, so $\mathbb{Q}(\sqrt{-3})$ is the splitting field of x^2+3 .

If $\omega = \frac{-1+\sqrt{-3}}{2}$ then ω, ω^2 are the roots of x^2+x+1 so $F(\omega)$ is the splitting field of x^2+x+1 (Why?).

Now $\sqrt{-3}=2\omega+1 \in \mathbb{Q}(\omega)$ gives $\mathbb{Q}(\sqrt{-3}) \subseteq \mathbb{Q}(\omega)$

Again $\omega = -\frac{1}{2} + \frac{\sqrt{-3}}{2} \in \mathbb{Q}(\sqrt{-3})$ gives $\mathbb{Q}(\omega) \subseteq \mathbb{Q}(\sqrt{-3})$. So that $\mathbb{Q}(\omega) = \mathbb{Q}(\sqrt{-3})$.

Hence x^2+3 and x^2+x+1 have same splitting field. Here notice that $[\mathbb{Q}(\omega) : \mathbb{Q}] = \text{degree of } x^2+x+1 = 2$.

EXAMPLE 12. Consider the polynomial x^3-2 over $\mathbb{Q}$. Its roots in $\mathbb{C}$ are $\alpha = \sqrt[3]{2}$, $\omega\alpha$, $\omega^2\alpha$ where ω is imaginary cube root of unity. None of these roots is a rational number, so x^3-2 is irreducible over $\mathbb{Q}$. Thus if E is the splitting field of x^3-2 over $\mathbb{Q}$ then $[E : \mathbb{Q}] \leq 3! = 6$. Now $[\mathbb{Q}(\alpha) : \mathbb{Q}] = 3$, $[\mathbb{Q}(\omega) : \mathbb{Q}] = 2$ as $\mathbb{Q}(\omega)$ and $\mathbb{Q}(\alpha)$ are both subfields of E , we get $3 \mid [E : \mathbb{Q}]$ and $2 \mid [E : \mathbb{Q}]$. Hence $6 \mid [E : \mathbb{Q}]$. Consequently $[E : \mathbb{Q}] = 6 = 3!$

EXAMPLE 13. Consider the polynomial x^3-3x+1 over $\mathbb{Q}$. It has no rational roots, so it is irreducible over $\mathbb{Q}$. Its roots in $\mathbb{R}$ are $-2 \cos \frac{\pi}{9}$, $2 \cos \frac{2\pi}{9}$, $2 \cos \frac{4\pi}{9}$. Let $\alpha = -2 \cos \frac{\pi}{9}$. Consider $\mathbb{Q}(\alpha) \subseteq E$, the splitting field of x^3-3x+1 .

Now $2 \cos \frac{2\pi}{9} = 4 \cos^2 \frac{\pi}{9} - 2 = \alpha^2 - 2 \in \mathbb{Q}(\alpha)$. $2 \cos \frac{4\pi}{9} = 4 \cos^2 \frac{2\pi}{9} - 2 = (\alpha^2 - 2)^2 - 2 \in \mathbb{Q}(\alpha)$. So all roots of x^3-3x+1 are in $\mathbb{Q}(\alpha)$. Thus $E \subseteq \mathbb{Q}(\alpha)$. Hence $E = \mathbb{Q}(\alpha)$. Further $[E : \mathbb{Q}] = [\mathbb{Q}(\alpha) : \mathbb{Q}] = \text{degree of minimal polynomial of } \alpha \text{ over } \mathbb{Q} = \deg(x^3-3x+1) = 3$. In this case $[E : \mathbb{Q}] = 3 < 3!$

EXAMPLE 15. Consider the polynomial x^4-1 over $\mathbb{Q}$. Its roots

are 1, -1, i , $-i$ where $i = \sqrt{-1}$. So the splitting field of $x^4 - 1$ over $\mathbb{Q}$ is $\mathbb{Q}(i)$ and $[\mathbb{Q}(i) : \mathbb{Q}] = 2$.

EXAMPLE 15. *Definition:* A complex number α is said to be an algebraic number if it is algebraic over $\mathbb{Q}$, the field of rational numbers. Since algebraic numbers form a field (Theorem 13.23), let A be the field of algebraic numbers. By definition, A is an algebraic extension of $\mathbb{Q}$. We claim that A is not a finite field extension of $\mathbb{Q}$. If it were, suppose $[A : \mathbb{Q}] = n$. Consider the polynomial $x^{n+1} - 3$ over $\mathbb{Q}$. By Eisenstein's Criterion, this is irreducible over $\mathbb{Q}$. Let α be a root of $x^{n+1} - 3$ then α is algebraic over $\mathbb{Q}$, so $\alpha \in A$. Now as $x^{n+1} - 3$ is irreducible over $\mathbb{Q}$, $[\mathbb{Q}(\alpha) : \mathbb{Q}] = n+1$, but $\alpha \in A$ implies that $\mathbb{Q}(\alpha) \subseteq A$. This is absurd as $\mathbb{Q}(\alpha)$ is a subspace of A over $\mathbb{Q}$ and $[A : \mathbb{Q}] = n$. So our assumption is wrong. Consequently A is an infinite field extension of $\mathbb{Q}$.

The above example shows that an algebraic extension need not be finite while every finite extension is an algebraic extension (Theorem 13.21).

Now we discuss the relation between roots and coefficients of a polynomial.

Theorem 13.37. If $f(x) = \alpha_0 + \alpha_1 x + \dots + \alpha_n x^n \in F[x]$ is a polynomial of degree $n \geq 1$ with roots $a_1, a_2, \dots, a_n$ in some extension

$$E \text{ of } F \text{ then } s_1 = \sum_i a_i = -\frac{\alpha_{n-1}}{\alpha_n}, s_2 = \sum_{i>j} a_i a_j = \frac{\alpha_{n-2}}{\alpha_n}, s_3 = \sum_{i>j>k} a_i a_j a_k \\ = -\frac{\alpha_{n-3}}{\alpha_n} \text{ etc., } \dots, s_n = a_1 a_2 \dots a_n = (-1)^n \frac{\alpha_0}{\alpha_n}.$$

Proof: Since $f(x)$ is of degree n , $\alpha_n \neq 0$, so $\alpha_n^{-1} = \frac{1}{\alpha_n} \in F$; further as $a_1, a_2, \dots, a_n$ are all the roots of $f(x)$.

$$f(x) = \alpha_n (x - a_1)(x - a_2) \dots (x - a_n).$$

This gives

$$\alpha_0 + \alpha_1 x + \alpha_2 x^2 + \dots + \alpha_{n-2} x^{n-2} + \alpha_{n-1} x^{n-1} + \alpha_n x^n \\ = \alpha_n [x^n - s_1 x^{n-1} + s_2 x^{n-2} + \dots + (-1)^n s_n]$$

where $s_1 = \sum_i a_i, s_2 = \sum_{i>j} a_i a_j, s_3 = \sum_{i>j>k} a_i a_j a_k$ and so on till

$$s_n = a_1 a_2 \dots a_n$$

Now as x is transcendental over F , we get comparing the coefficients of like powers of x ,

$$-\alpha_n s_1 = \alpha_{n-1}, \alpha_n s_2 = \alpha_{n-2}, -\alpha_n s_3 = \alpha_{n-3}, \dots$$

and so on, lastly $(-1)^n \alpha_n s_n = \alpha_0$.

Since α_n is a unit in F , we get

$$s_1 = -\frac{\alpha_{n-1}}{\alpha_n}, s_2 = \frac{\alpha_{n-2}}{\alpha_n}, s_3 = -\frac{\alpha_{n-3}}{\alpha_n}, \dots$$

$$\dots, s_n = (-1)^n \frac{\alpha_0}{\alpha_n}. \blacksquare$$

WORKED-OUT EXERCISES

Note. Throughout the following worked-out exercises, K is a field extension of F .

Exercise 1. $a, b \in K$ are algebraic over F of degrees m and n respectively, where m and n are relatively prime; prove that $F(a, b)$ is of degree mn over F .

Solution. We are given that $[F(a) : F] = m$ and $[F(b) : F] = n$. Let $x^m + \alpha_1 x^{m-1} + \dots + \alpha_m$ be the minimal polynomial of a over F . This is also a polynomial over $F(b)[x]$. Thus a satisfies a polynomial of degree m over $F(b)$.

Consequently $[F(b)(a) : F(b)] \leq m$, i.e., $[F(a, b) : F(b)] \leq m$. Now $[F(a, b) : F] = [F(a, b) : F(b)] [F(b) : F] \leq mn$. Again $[F(a, b) : F] = [F(a, b) : F(a)] [F(a) : F] \Rightarrow m \mid [F(a, b) : F]$. Similarly $[F(a, b) : F] = [F(a, b) : F(b)] [F(b) : F] \Rightarrow n \mid [F(a, b) : F]$. As $(m, n) = 1$, we get that $mn \mid [F(a, b) : F]$. So $mn \leq [F(a, b) : F]$. .. (1)

(1) and (2) gives the required result.

Exercise 2. Prove that $a \in K$ is algebraic over F of an odd degree then $F(a) = F(a')$. .. (2)

Solution. Clearly $a^n \in F(a) \Rightarrow F(a^2) \subseteq F(a)$.

Let $p(x) = x^{2n+1} + \alpha_0 x^{2n} + \dots + \alpha_{2n}$ be the minimal polynomial of a over F . Then $a^{2n+1} + \alpha_0 a^{2n} + \dots + \alpha_{2n} = 0$.

$\Rightarrow a(a^{2n} + \alpha_1 a^{2n-2} + \dots + \alpha_{2n-1}) + (\alpha_0 a^{2n} + \alpha_2 a^{2n-2} + \dots + \alpha_{2n}) = 0$.

Now as $[F(a) : F] = 2n+1$, $a^{2n} + \alpha_1 a^{2n-2} + \dots + \alpha_{2n-1} \neq 0$; so

$a = -(a^{2n} + \alpha_1 a^{2n-2} + \dots + \alpha_{2n-1})^{-1} (\alpha_0 a^{2n} + \alpha_2 a^{2n-2} + \dots + \alpha_{2n}) \in F(a^2)$.

Thus we see that $F(a) \subseteq F(a^2)$. Hence $F(a) = F(a^2)$.

Exercise 3. Find the degree of splitting field of $x^5 - 3x^3 + x^2 - 3$ over $\mathbb{Q}$.

Solution. Now $x^5 - 3x^3 + x^2 - 3 = (x^2 - 3)(x^3 + 1)$ gives that roots of given polynomial in $\mathbb{C}$ are $\pm\sqrt{3}, -1, \frac{1+i\sqrt{3}}{2}$. So the splitting field of $x^5 - 3x^3 + x^2 - 3$ over $\mathbb{Q}$ is $K = \mathbb{Q} \left(\pm\sqrt{3}, -1, \frac{1+i\sqrt{3}}{2} \right)$.

We claim that $K = \mathbb{Q}(\sqrt{3}, i)$. Now $\frac{1+i\sqrt{3}}{2} \in K$

$\Rightarrow \left(\frac{1+i\sqrt{3}}{2} \right) - \left(\frac{1-i\sqrt{3}}{2} \right) = i\sqrt{3} \in K$. Also $\sqrt{3} \in K$ so $i \in K$.

Thus $\mathbb{Q}(\sqrt{3}, i) \subseteq K$. Again $\sqrt{3} \in \mathbb{Q}(\sqrt{3}, i) \Rightarrow \pm i\sqrt{3} \in \mathbb{Q}(\sqrt{3}, i)$. As

a consequence $\frac{1 \pm i\sqrt{3}}{2} \in \mathbb{Q}(\sqrt{3}, i)$. Thus $K \subseteq \mathbb{Q}(\sqrt{3}, i)$. Hence our claim is established.

Now $[K : \mathbb{Q}] = [K : \mathbb{Q}(i)][\mathbb{Q}(i) : \mathbb{Q}] \leq 2 \cdot 2 = 4$, as minimal polynomial of i over $\mathbb{Q}$ is $x^2 + 1$. Also $[K : \mathbb{Q}] = [K : \mathbb{Q}(\sqrt{3})][\mathbb{Q}(\sqrt{3}) : \mathbb{Q}]$ shows that 2 is a factor of $[K : \mathbb{Q}]$. So we get that either $[K : \mathbb{Q}] = 2$ or 4. In case $[K : \mathbb{Q}] = 2$, we shall get $[K : \mathbb{Q}(\sqrt{3})] = 1$, i.e., $K = \mathbb{Q}(\sqrt{3})$ which is absurd as $i \in K$ but $i \notin \mathbb{Q}(\sqrt{3})$. Hence $[K : \mathbb{Q}] = 4$.

Exercise 4. Find the splitting field of $x^p - 1$ over $\mathbb{Q}$, p being a prime number.

Solution. Roots of $x^p - 1$ in $\mathbb{C}$ are $1, e^{\frac{2\pi i}{p}}, e^{\frac{4\pi i}{p}}, \dots, e^{\frac{2(p-1)\pi i}{p}}$.

If $\xi = e^{\frac{2\pi i}{p}}$ then the splitting field of $x^p - 1$ over $\mathbb{Q}$ is $K = \mathbb{Q}(1, \xi, \xi^2, \dots, \xi^{p-1}) = \mathbb{Q}(\xi)$. (Prove!).

Now ξ is a root of $\frac{x^p - 1}{x - 1} = x^{p-1} + x^{p-2} + \dots + 1$. But this is minimal polynomial of ξ over $\mathbb{Q}$. (See Example 30, Chapter 10.) Hence $[\mathbb{Q}(\xi) : \mathbb{Q}] = p - 1$, i.e., $[K : \mathbb{Q}] = p - 1$.

Exercise 5. A field F is said to be algebraically closed if every non-constant polynomial in $F(x)$ splits completely (i.e., has all its roots) in F . Prove that an algebraically closed field can not be finite.

Solution. Suppose F is a finite algebraically closed field. Let $F = \{a_1, a_2, \dots, a_n\}$. Consider the polynomial $f(x) = 1 + \prod_{i=1}^n (x - a_i) \in F[x]$. It must have all its roots in F as F is algebraically closed. But none of the elements a_i is a root of $f(x)$. So we get a contradiction. Hence F can not be finite.

Exercise 6. Let F be a field of characteristic different from 2 and let $\alpha, \beta \in F$ be such that $x^2 - \alpha$ and $x^2 - \beta$ are irreducible over F . If K is the splitting field of $f(x) = (x^2 - \alpha)(x^2 - \beta)$ over F then show that $[K : F]$ is 2 or 4 according as $x^2 - \alpha\beta$ is reducible or irreducible over F .

Solution. Roots of $f(x)$ in any extension of F are $\pm\sqrt{\alpha}, \pm\sqrt{\beta}$. So $K = F(\pm\sqrt{\alpha}, \pm\sqrt{\beta}) = F(\sqrt{\alpha}, \sqrt{\beta})$. Now $[F(\sqrt{\beta}) : F] = 2 \Rightarrow [F(\sqrt{\alpha}, \sqrt{\beta}) : F(\sqrt{\alpha})] \leq 2$. In case $x^2 - \alpha\beta$ is reducible F , we have $\sqrt{\alpha\beta} \in F$.

This in turn implies that $\sqrt{\alpha\beta} \in F(\sqrt{\alpha})$. But $\sqrt{\alpha} \in F(\sqrt{\alpha})$ so $\sqrt{\beta} \in F(\sqrt{\alpha})$. As a consequence $[F(\sqrt{\alpha}, \sqrt{\beta}) : F(\sqrt{\alpha})] = 1$ and thus $[K : F] = [F(\sqrt{\alpha}, \sqrt{\beta}) : F(\sqrt{\alpha})][F(\sqrt{\alpha}) : F] = 2$.

Let now $x^2 - \alpha\beta$ be irreducible over F . As $\sqrt{\alpha\beta}$ is a root of $x^2 - \alpha\beta$ over $F(\sqrt{\alpha})$, we get $[F(\sqrt{\alpha\beta}, \sqrt{\alpha}) : F(\sqrt{\alpha})]$

≤ 2 . If $[F(\sqrt{\alpha\beta}, \sqrt{\alpha}) : F(\sqrt{\alpha})] = 1$, we shall get $\sqrt{\alpha\beta} \in F(\sqrt{\alpha})$
 $\Rightarrow \sqrt{\alpha\beta} = a + b\sqrt{\alpha}$, $a, b \in F \Rightarrow \alpha\beta = a^2 + ab^2 + 2ab\sqrt{\alpha} \Rightarrow 2ab\sqrt{\alpha} = \alpha\beta - a^2 - ab^2 \in F$
 $\Rightarrow \sqrt{\alpha} \in F$ since $a \neq 0$, $b \neq 0$ and $\text{char } F \neq 2$.

[Note that $a=0 \Rightarrow \sqrt{\beta} = b \in F \Rightarrow x^2 - \beta$ is reducible over F ! While if $b=0$ then we shall get $\sqrt{\alpha\beta} = a \Rightarrow x^2 - \alpha\beta$ is reducible over F !]

This in turn implies that $x^2 - \alpha$ is reducible over F , a contradiction. Hence $[F(\sqrt{\alpha\beta}, \sqrt{\alpha}) : F(\sqrt{\alpha})] = 2$. But $F(\sqrt{\alpha\beta}, \sqrt{\alpha}) \subseteq F(\sqrt{\alpha}, \sqrt{\beta})$, so $[F(\sqrt{\alpha}, \sqrt{\beta}) : F(\sqrt{\alpha})] \geq 2$. Hence $[F(\sqrt{\alpha}, \sqrt{\beta}) : F(\sqrt{\alpha})] = 2$. As a consequence we get $[K : F] = 4$.

Exercise 7. Show that $\sqrt{2} + \sqrt[3]{5}$ is algebraic over $\mathbb{Q}$ of degree 6.

Solution. Put $\alpha = \sqrt{2} + \sqrt[3]{5}$. This yields that $\alpha - \sqrt{2} = \sqrt[3]{5} \Rightarrow (\alpha - \sqrt{2})^3 = 5 \Rightarrow \alpha^3 + 6\alpha - 5 = (3\alpha^2 + 2)\sqrt{2} \Rightarrow (\alpha^3 + 6\alpha - 5)^2 = 2(3\alpha^2 + 2)^2$
 $\Rightarrow \alpha$ is algebraic over $\mathbb{Q}$ of degree less than or equal to 6. To prove the second assertion we claim that $\mathbb{Q}(\sqrt{2} + \sqrt[3]{5}) = \mathbb{Q}(\sqrt{2}, \sqrt[3]{5})$. Then invoking the result of exercise 1, we shall get the second part.

Now $\sqrt{2}, \sqrt[3]{5} \in \mathbb{Q}(\sqrt{2}, \sqrt[3]{5}) \Rightarrow \sqrt{2} + \sqrt[3]{5} \in \mathbb{Q}(\sqrt{2}, \sqrt[3]{5})$
 $\Rightarrow \mathbb{Q}(\sqrt{2} + \sqrt[3]{5}) \subseteq \mathbb{Q}(\sqrt{2}, \sqrt[3]{5})$

Further $\sqrt{2} + \sqrt[3]{5} \in \mathbb{Q}(\sqrt{2} + \sqrt[3]{5}) \Rightarrow (\sqrt{2} + \sqrt[3]{5})^2 \in \mathbb{Q}(\sqrt{2} + \sqrt[3]{5})$... (1)
i.e., $2 + 2\sqrt{2}\sqrt[3]{5} + \sqrt[3]{25} \in \mathbb{Q}(\sqrt{2} + \sqrt[3]{5}) \Rightarrow \sqrt[3]{5}(\sqrt{2} + \sqrt{2} + \sqrt[3]{5}) \in \mathbb{Q}(\sqrt{2} + \sqrt[3]{5})$
 $\Rightarrow 5[(\sqrt{2} + (\sqrt{2} + \sqrt[3]{5}))^3] \in \mathbb{Q}(\sqrt{2} + \sqrt[3]{5})$.

Now we have $5(\sqrt{2} + \alpha)^3 \in \mathbb{Q}(\alpha) \Rightarrow (\sqrt{2} + \alpha)^3 \in \mathbb{Q}(\alpha)$
 $\Rightarrow 2\sqrt{2} + 6\alpha + 3\sqrt{2}\alpha^2 + \alpha^3 \in \mathbb{Q}(\alpha) \Rightarrow \sqrt{2}(2 + 3\alpha^2) \in \mathbb{Q}(\alpha)$ as $6\alpha + \alpha^3 \in \mathbb{Q}(\alpha)$, so $\sqrt{2} \in \mathbb{Q}(\alpha)$.

This shows that $\sqrt[3]{5} = \alpha - \sqrt{2} \in \mathbb{Q}(\alpha)$. Hence $\sqrt{2}, \sqrt[3]{5} \in \mathbb{Q}(\alpha)$
 $\Rightarrow \mathbb{Q}(\sqrt{2}, \sqrt[3]{5}) \subseteq \mathbb{Q}(\sqrt{2} + \sqrt[3]{5})$.

(1) and (2) imply that $\mathbb{Q}(\sqrt{2}, \sqrt[3]{5}) = \mathbb{Q}(\sqrt{2} + \sqrt[3]{5})$ (2)

PROBLEMS

Note : In the following problems F stands for a field and K for its field extension, unless stated otherwise.

1. If $[K : F] = 1$ then prove that $K = F$.
2. What is the difference between " K is a finite extension of F " and " K is finitely generated". Give example to illustrate your answer.
[Hint. If x is transcendental over F then $K = F(x)$ is finitely generated over F but not finite extension of K .]

3. For any two elements a, b in K , prove that $F(a, b) = F(b, a)$.
4. Prove that a cubic over a field F is irreducible over F if and only if it has no root in F .

[Hint. If $ax^3 + bx^2 + cx + d$ is reducible then $ax^3 + bx^2 + cx + d = a(x + \alpha)(x^2 + \beta x + \gamma)$ for some $\alpha, \beta, \gamma \in F$.]

5. Let x^2+ax+b be a quadratic over F . Prove that if α is a root of this quadratic in any field extension of F , then $\beta=-(a+\alpha)$ is also a root of the quadratic and the splitting field of this quadratic over F is given by $F(\alpha)$.

6. Find a basis and the degree of $\mathbb{Q}(\sqrt{2}, \sqrt{3})$ over F .

[Ans. $\{1, \sqrt{2}, \sqrt{3}, \sqrt{6}\}$, 4]

7. Prove that $F(\sqrt{2}+\sqrt{3})=F(\sqrt{2}, \sqrt{3})$.

8. If F is a subfield of a domain D such that each element of D is algebraic over F , show that D is a field. Deduce that if K is any algebraic extension of F , then any subdomain D of K containing F is a subfield of K .

9. For any two subfields M and N of a field, define $MN=\{\sum x_i y_i \mid x_i \in M, y_i \in N\}$. Prove the following:

(i) MN is a subring of K .

(ii) If K is an algebraic extension of F and M, N contain F then MN is the subfield of K containing F such that $MN=\langle M \cup N \rangle$. In this case if M and N are finite-dimensional over F , deduce that if $[MN:F]=[M:F][N:F]$ then $M \cap N=F$. Prove that the converse holds if one of $[M:F]$ and $[N:F]$ is two. Give an example to show that in general the converse does not hold.

[Hint: Take $M=\mathbb{Q}(\sqrt[3]{2})$, $N=\mathbb{Q}(\omega\sqrt[3]{2})$, $F=\mathbb{Q}$, where ω is imaginary cube root of unity.]

10. In $\mathbb{Q}(\sqrt{2})$ express the following elements as polynomials in $\sqrt{2}$.

$$(i) \frac{1}{2+\sqrt{2}} \quad (ii) \frac{3+\sqrt{2}}{5+\sqrt{8}} \quad (iii) \frac{1}{(3+\sqrt{2})}$$

$$\left[\text{Ans. (i) } 1-\frac{1}{2}\sqrt{2} \quad (ii) \frac{11}{17}-\frac{1}{17}\sqrt{2} \quad (iii) \frac{11}{49}-\frac{6}{49}\sqrt{2} \right]$$

11. In $\mathbb{Q}(\sqrt[3]{2})$ express the following elements as polynomials in $\sqrt[3]{2}$.

$$(i) \frac{3}{1+\sqrt[3]{2}-\sqrt[3]{4}} \quad (ii) (1+\sqrt[3]{2}+6\sqrt[3]{4})^{-1} \quad (iii) \frac{1}{2-\sqrt[3]{4}}$$

$$\left[\text{Ans. (i) } \frac{9}{5}+\frac{3}{5}\sqrt[3]{2}+\frac{6}{5}\sqrt[3]{4} \quad (ii) -\frac{11}{831}+\frac{71}{831}\sqrt[3]{2}-\frac{5}{831}\sqrt[3]{4} \right.$$

$$\left. (iii) 1+\frac{1}{2}\sqrt[3]{2}+\frac{1}{2}\sqrt[3]{4} \right]$$

12. Let x^3+ax^2+bx+c be any irreducible cubic over $\mathbb{Q}$ and let E be its splitting field. Show that all the roots of cubic are real whenever $[E:\mathbb{Q}]=3$.

13. Show that the splitting field of x^6-1 over $\mathbb{Q}$ is same as that of x^4+x^2+1 over $\mathbb{Q}$ and $[E:\mathbb{Q}]=2$.

14. Find the splitting fields and their degrees, of the following polynomials over $\mathbf{Q}$.

- (i) x^4+1 (ii) x^6+1 (iii) x^4-2 (iv) x^5-1 .

[Ans. (i) $\mathbf{Q}(\sqrt{2}, i)$, 4. (ii) $\mathbf{Q}(\sqrt[3]{3}, i)$, 4. (iii) $\mathbf{Q}(2^{1/4}, i)$, 8
(iv) $\mathbf{Q}(e^{2\pi i/5})$, 4.]

15. Find necessary and sufficient conditions for an irreducible cubic x^3+ax+b over $\mathbf{Q}$ to have a splitting field of degree 3.

[Hint. Let $D = -4a^3 - 27b^2$, then the degree of splitting field is 3 if and only if D is a perfect square.]

16. Let $f(x)$ be any polynomial over F and σ be any automorphism of a field $K \supseteq F$ such that σ leaves every element of F fixed. Show that for any root α of $f(x)$ in K , $\sigma(\alpha)$ is also a root of $f(x)$.

17. Let m be any positive integer, which is not a perfect square. Show that the mapping $\alpha + \beta\sqrt{m} \rightarrow \alpha - \beta\sqrt{m}$, $\alpha, \beta \in \mathbf{Q}$ is an automorphism of $\mathbf{Q}(\sqrt{m})$. Hence or otherwise show that for any two rational numbers α, β with $\beta \neq 0$, if $p(x)$ is the minimal polynomial of $\alpha + \beta\sqrt{m}$ over $\mathbf{Q}$, then $p(x)$ is also the minimal polynomial of $\alpha - \beta\sqrt{m}$ over $\mathbf{Q}$.

18. Using problem 16, show that if a complex number $a+ib$ is a root of a polynomial $f(x)$ over $\mathbf{R}$, the field of real numbers, then $a-ib$ is also a root of $f(x)$.

[Hint. $z \rightarrow \bar{z}$, $z \in \mathbf{C}$, is an automorphism of $\mathbf{R}(i)$ fixing $\mathbf{R}$ pointwise.]

19. Let F be any field and D , a division ring such that F is contained in the centre of D . Give an example to show that a polynomial $f(x)$ of degree $n > 1$ over F may have more than n roots in D .

[Hint. Let D be the division ring of real quaternions, show that x^2+1 has infinitely many roots in D .]

20. Let u and $v \in K$ be algebraic over F of degree m and n respectively. Show that u has degree m over $K(v)$ if and only if v has degree n over $K(u)$ and that both statements hold if m and n are relatively prime.

21. We say that a field is *absolutely algebraic* if it is algebraic over its prime subfield. Show that an absolutely algebraic field of finite characteristic has no subdomain which is not a field.

22. Let $F_1 \supseteq F_2 \supseteq \dots \supseteq F_n$ be a finite chain of fields such that each

$[F_i : F_{i+1}]$ is finite then show that $[F_1 : F_n] = \prod_{i=1}^{n-1} [F_i : F_{i+1}]$.

23. *Definition.* A complex number is said to be an *algebraic integer* if it satisfies a monic polynomial in $\mathbf{Z}[x]$. i is an algebraic integer as it satisfies x^2+1 .

Prove the following:

(i) If a rational number is an algebraic integer then it must be an integer.

(ii) If α is an algebraic number then there exists a positive integer n such that $n\alpha$ is algebraic integer.

(iii) Algebraic integers form a subring of $\mathbb{C}$.

24. Prove that for every integer m both $\cos m^\circ$ and $\sin m^\circ$ are algebraic numbers. [Hint: $\left(e^{\frac{\pi mi}{180}}\right)^{180} = \pm 1$].

3. Separable and Inseparable Extensions

Let F be any field and $f(x)$, a non-constant polynomial over F . In the last section we introduced the concept of multiplicity of a root of $f(x)$ in any field extension. A root a of $f(x)$ is said to be a *multiple root* of $f(x)$, if a is a root of multiplicity greater than or equal to two. In this section we intend to discuss the problem of existence of multiple roots of polynomials. For this we introduce the concept of a derivative.

Definition 13.38. If $f(x) = \alpha_0 + \alpha_1 x + \alpha_2 x^2 + \dots + \alpha_n x^n$ is a polynomial in $F[x]$, then the *derivative* of $f(x)$ which is denoted by $f'(x)$, is defined as $f'(x) = \alpha_1 + 2\alpha_2 x + \dots + n\alpha_n x^{n-1}$.

Thus in particular, if $f(x) = \alpha_0$ then $f'(x) = 0$, i.e. derivative of any constant polynomial is zero. Further it is clear from the definition that if $f(x)$ is a polynomial of degree n , then either $f'(x) = 0$ or the degree of $f(x)$ is at the most $n-1$. To illustrate this let F be of characteristic p and $g(x) = x^p$. Then $g'(x) = px^{p-1} = 0$, since the characteristic of $F[x]$ is also p . However if we take F to be of characteristic 0, and $\deg f(x) = n \geq 1$ then $\alpha_n \neq 0$ gives $n\alpha_n \neq 0$. So that $\deg f'(x) = n-1$.

Now the following lemma is an immediate consequence of the definition of derivative.

Lemma 13.39. Let $f(x), g(x) \in F[x]$, $\alpha \in F$ then

(i) $[f(x) + g(x)]' = f'(x) + g'(x)$

(ii) $[\alpha f(x)]' = f'(x)$

(iii) $[f(x)g(x)]' = f'(x)g(x) + f(x)g'(x)$. ■

Lemma 13.40. Let $f(x) \in F[x]$ be a non-constant polynomial, then an element α of a field extension K of F is a multiple root of $f(x)$ if and if α is a common root of $f(x)$ and $f'(x)$.

Proof: Let α be a root of $f(x)$ of multiplicity m . Then $f(x) = (x - \alpha)^m g(x)$, $g(x) \in K[x]$ and $g(\alpha) \neq 0$.

We get $f'(x) = m(x - \alpha)^{m-1} g(x) + (x - \alpha)^m g'(x)$.

Let α be a multiple root of $f(x)$ then $m \geq 2$, so

$$f'(\alpha) = m(\alpha - \alpha)^{m-1}g(\alpha) + (\alpha - \alpha)^mg'(\alpha) = 0 \text{ since } m-1 > 0.$$

That means α is a common root of $f(x)$ and $f'(x)$.

Notice that if α is not a multiple root then $m=1$ and $f'(x)=g(x) + (x-\alpha)g'(x)$. In this case $f'(\alpha)=g(\alpha) \neq 0$. This proves the lemma. ■

Lemma 13.41. Let $f(x)$ be an irreducible polynomial over F . Then $f(x)$ has a multiple root in some field extension [definitely in the splitting field of $f(x)$ over F] if and only if $f'(x)=0$.

Proof: Let $f(x)$ have a multiple root a in a field extension K of F . Then a is also a root of $f'(x)$, (Lemma 13.40). Now as $f(x)$ is an irreducible polynomial over F satisfied by a , so that for any non-zero polynomial $g(x)$, $g(a)=0$ implies $f(x)$ divides $g(x)$. In particular $f(x) \mid f'(x)$. If $f'(x) \neq 0$ then we know that $\deg f'(x) < \deg f(x)$. Consequently $f(x)$ cannot divide $f'(x)$. Hence $f'(x)=0$.

Conversely let $f'(x)=0$. Let K be the splitting field of $f(x)$ over F . Let $\deg f(x)=n$.

Then $f(x) = \alpha(x-a_1)(x-a_2) \dots (x-a_n)$ where $a_i (i=1, 2, \dots, n)$ are the n roots of $f(x)$ in K and $\alpha (\neq 0)$ is in F .

Now $f'(x) = \alpha \sum_{i=1}^n (x-a_1)(x-a_2) \dots \widehat{(x-a_i)} \dots (x-a_n)$ where for each

i , $(x-a_1)(x-a_2) \dots \widehat{(x-a_i)} \dots (x-a_n)$ denotes the product of all $(x-a_j)$'s with $j \neq i$. Then $f'(a_i) = (a_i-a_1)(a_i-a_2) \dots (a_i-a_{i-1})(a_i-a_{i+1}) \dots (a_i-a_n) \neq 0$ unless for some $j \neq i$, $a_j = a_i$. So that $f'(x)=0$ cannot hold unless some one of the roots, say a_i , is a multiple root of $f(x)$. Hence $f(x)$ has a multiple root in K . ■

Corollary 13.42. No irreducible polynomial over a field of characteristic zero has a multiple root in any field extension.

Proof: Let $f(x)$ be an irreducible polynomial over a field F of characteristic 0. Let $\deg f(x)$ be $n \geq 1$, in this case $\deg f'(x) = n-1 \geq 0$ (Remark following definition 13.38).

Thus $f'(x) \neq 0$. Hence by the above theorem $f(x)$ can not have a multiple root in any field extension of F . ■

EXAMPLE 16. Let $F = \mathbb{Z}_2(t)$, where t is an indeterminate over $\mathbb{Z}_2$, the field of integers modulo 2. Consider the polynomial $f(x) = x^2 - t$ over F . Let K be the splitting field of $f(x)$ and a_1, a_2 be the two roots of $f(x)$ in K . Then $x^2 - t = (x-a_1)(x-a_2)$. Comparing coefficients, we get $a_1 = -a_2$. However, characteristic of F is 2, gives that $-a_2 = a_2$. So $a_1 = a_2$ and $f(x) = (x-a_1)^2$, showing thereby that a_1 is a multiple root of $f(x)$. It can be seen that $f(x)$ is irreducible over F . (See problem 6 at the end of this section). Further $f'(x) = 2x = 0$.

Definition 13.43. An irreducible polynomial $f(x) \in F[x]$ of degree n is said to be separable if it has n distinct roots in its splitting field

otherwise it is said to be *inseparable*. In general a non-constant polynomial $g(x) \in F(x)$ is said to be *separable* if all the irreducible factors of $g(x)$ over F are separable otherwise $g(x)$ is said to be *inseparable*.

It follows from Lemma 13.41, that an irreducible polynomial $f(x)$ is separable if and only if $f'(x) \neq 0$. Corollary 13.42 gives that every non-zero polynomial over a field of characteristic zero is separable. Example 16 is an example of an irreducible inseparable polynomial. Consider $f(x) = (x-1)^n(x^2+x+1)$ over $\mathbb{Q}$. It has 1 as a repeated root. However its irreducible factors are $x-1$ and x^2+x+1 both of them are separable. So by definition $f(x)$ is separable.

Lemma 13.44. *An irreducible polynomial $f(x)$ over a field F of characteristic $p > 0$ is inseparable if and only if $f(x) \in F[x^p]$, (i.e., $f(x)$ is a polynomial in x^p).*

Proof: Let $f(x) = \alpha_0 + \alpha_1 x + \alpha_2 x^2 + \dots + \alpha_n x^n$ with $\alpha_n \neq 0$. Then $f'(x) = \alpha_1 + 2\alpha_2 x + \dots + n\alpha_n x^{n-1}$. Now $f(x)$ is inseparable $\Leftrightarrow f'(x) = 0$. But $f'(x) = 0 \Leftrightarrow \alpha_1 = 2\alpha_2 = 3\alpha_3 = \dots = n\alpha_n = 0$. For any k , $1 \leq k \leq n$, $k\alpha_k = 0 \Leftrightarrow p \mid k$ or $\alpha_k = 0$. Thus whenever $f'(x) = 0$ we see that if for any k , $\alpha_k \neq 0$ then $p \mid k$ i.e., $k = k_1 p$ for some integer $k_1 \geq 1$. That means, in $f(x)$ if any term $\alpha_k x^k$ has $\alpha_k \neq 0$, then it is of the type $\alpha_{k_1 p} x^{p k_1} = \alpha_{k_1 p} (x^p)^{k_1}$. So that $f(x) = \beta_0 + \beta_1 x^p + \beta_2 x^{2p} + \dots + \beta_m x^{mp}$ for some positive integer m . Thus $f(x) \in F[x^p]$. Hence the result follows. ■

Definition 13.45. If an element a of a field extension K of F is algebraic over F , then a is said to be *separable* (inseparable) over F , if the minimal polynomial of a over F is separable (inseparable).

An algebraic extension K of a field F is said to be a *separable extension*, if every element of K is separable over F otherwise K is said to be an *inseparable extension*.

As observed before every polynomial over a field of characteristic zero is separable, we see that every algebraic extension of a field of characteristic zero is a separable extension.

However if we take $F = \mathbb{Z}_2(t)$ and K be the splitting field of $x^2 - t$ (Example 16), then K is a finite extension of F as $[K:F] = 2$. But K is an inseparable extension as $x^2 - t$ has a repeated root in K . Notice that F is an infinite field of finite characteristic. We show that any algebraic extension of a finite field is separable. For this we establish first the following:

Lemma 13.46. *Let D be any integral domain of characteristic p , p prime number. Then the mapping $\sigma : D \rightarrow D$ given by $\sigma(a) = a^p$ is a monomorphism of D into D .*

Proof: For any $a, b \in D$

$$(a+b)^p = a^p + {}^pC_1 a^{p-1}b + {}^pC_2 a^{p-2}b^2 + \dots + {}^pC_p b^p$$

Now for each r with $1 \leq r \leq p-1$, $p \mid {}^pC_r$.

So that ${}^pC_r a^{p-r}b^r = 0$ for all $r=1, 2, \dots, p-1$.

Hence $(a+b)^p = a^p + b^p$. Also the commutativity of multiplication gives $(ab)^p = a^p b^p$. So that

$$\sigma(a+b) = (a+b)^p = a^p + b^p = \sigma(a) + \sigma(b)$$

$$\sigma(ab) = (ab)^p = a^p b^p = \sigma(a) \sigma(b).$$

Thus σ is an endomorphism. Further $a \in \text{Ker } \sigma \Rightarrow \sigma(a) = 0$

$\Rightarrow a^p = 0 \Rightarrow a = 0$. Hence $\text{Ker } \sigma = (0)$ and σ is a monomorphism. ■

Corollary 13.47. If F is a finite field of characteristic p then $a \rightarrow a^p$ is an automorphism of F .

Proof: We have seen in the above lemma that the mapping $\sigma: F \rightarrow F$ given by $\sigma(a) = a^p \forall a \in F$ is a monomorphism. So $\sigma(F)$ has same number of elements as in F . However F is finite and $\sigma(F) \subseteq F$, we get $\sigma(F) = F$. Hence σ is an automorphism. ■

Theorem 13.48. Any algebraic extension of a finite field F is a separable extension.

Proof: Let $f(x)$ be any irreducible polynomial over F . Suppose $f(x)$ is inseparable over F so $f(x) \in F[x^p]$ (Lemma 13.44), $f(x) = \beta_0 + \beta_1 x^p + \beta_2 x^{2p} + \dots + \beta_m x^{mp}$ for some $\beta_i \in F$, $(0 \leq i \leq m)$. As $\alpha \rightarrow \alpha^p$, $\alpha \in F$ is an automorphism of F we can find $\alpha_i \in F$ such that $\beta_i = \alpha_i^p \forall i$.

$$\text{Consequently } f(x) = \alpha_0^p + \alpha_1^p x^p + \alpha_2^p x^{2p} + \dots + \alpha_m^p x^{mp}$$

$$= (\alpha_0 + \alpha_1 x + \alpha_2 x^2 + \dots + \alpha_m x^m)^p, \text{ (Lemma 13.46).}$$

This implies $f(x)$ is not irreducible. This is a contradiction. Hence $f(x)$ is separable. Thus in particular if K is any algebraic extension of F , the minimal polynomial of each element of K over F is separable. Hence by definition K is a separable extension of F . ■

Let us recall that a field extension K of F is said to be a simple extension of F if $K = K(a)$ for some $a \in K$ and a is known as a primitive element of K over F . A theorem of great significance in field theory is the Lagrange's theorem of primitive element which states that every finite separable field extension is a simple extension. The proof of this theorem requires some deep results, which we cannot give here. However we prove here a special case of this theorem.

Theorem 13.49. If F is of characteristic zero and two elements a and b in some field extension of F are algebraic over F , then $F(a, b) = F(c)$ for some $c \in F(a, b)$.

Proof: Let $f(x)$ and $g(x)$ be the minimal polynomials over F of a and b respectively and let m, n be their respective degrees. Let K be

$$a_{n-1})=F(b).$$

Thus $K=F(b, a_n)$ (Why?). Now Theorem 13.41 gives $K=F(c)$ for some $c \in F(b, a_n)$. This proves the theorem. ■

Definition 13.51. (Galois Field). A field having finite number of elements is called a *Galois field*.

EXAMPLE 17. Consider the ring $\mathbb{Z}_p$ of integers modulo $\frac{p}{p}$ being a prime number. $\mathbb{Z}_p$ has p elements namely $\bar{0}, \bar{1}, \bar{2}, \dots, \overline{p-1}$. Recall that $\mathbb{Z}_p - \{0\}$ is a group under multiplication (Cf. Example 7, Remark 2, § 1, Chapter 2). So each non-zero element of $\mathbb{Z}_p$ is invertible. Hence $\mathbb{Z}_p$ is a field with p elements.

EXAMPLE 18. The splitting field K of $x^2+x+\bar{1}$ over $\mathbb{Z}_2$ consists of $\bar{0}, \bar{1}, \alpha$ and $\bar{1}+\alpha$ where $\alpha^2+\alpha+\bar{1}=0$. K is Galois field having four elements.

EXAMPLE 19. Let K be the splitting field of $x^2+\bar{1}$ over $\mathbb{Z}_3$. It has 9 elements, viz. $\bar{0}, \bar{1}, \bar{2}, \alpha, \bar{2}\alpha, \bar{1}+\alpha, \bar{1}+\bar{2}\alpha, \bar{2}+\alpha, \bar{2}+\bar{2}\alpha$ where $\alpha^2+1=0$. K is a Galois field having 9 elements.

Theorem 13.52. Let F be a field having q elements. Then

- (i) $\text{Char } F$ is p for some prime number p .
- (ii) $q=p^n$ for some $n \in \mathbb{N}$.

Proof: Let e be the identity (multiplicative) of F . As F has q elements; $e, 2e, 3e, \dots, (q+1)e$ cannot be all distinct. There exist, therefore, integers $l, k, 1 \leq l < k \leq q+1$ such that $ke=le, i.e., (k-l)e=0$. Let p be the least positive integer satisfying $pe=0$. We claim that p is a prime number. Suppose $p=rs$ with $1 < r < p$ and $1 < s < p$. Then $pe=0 \Rightarrow pe^2=0 \Rightarrow rse^2=0 \Rightarrow (re)(se)=0 \Rightarrow re=0$ or $se=0$ as we are working in a field. But each of the above conclusions is against the choice of p . Hence p is a prime number. This proves that $\text{char } F=p$.

For second part we utilize the result of Theorem 13.5. The prime subfield P of F being isomorphic to $\mathbb{Z}/\langle p \rangle$ has p elements. Since F is a finite field, $[F:P]=n$ for some $n \in \mathbb{N}$. Let $\{a_1, a_2, \dots, a_n\}$ be a basis of F over P . Each element of F can be written uniquely as $\sum_{i=1}^n \lambda_i a_i$ with $\lambda_i \in P$. As each λ_i can be chosen in p ways, the total

number of elements of F is p^n . Hence $q=p^n$. ■

Theorem 13.52 merely states that number of elements in a finite field must be a power of a certain prime number. Thus there cannot exist a field with 6, 10 or 12 elements. In the other direction, we shall show that given any prime number p and a strictly positive integer n , there exists a field having p^n elements. The existence of such a field is guaranteed by a theorem of E.H. Moore. Firstly we prove a lemma.

Lemma 13.53. *If a field F has q elements, then F is a splitting field of $x^q - x$ over its prime subfield.*

Proof: Since non-zero elements of a field form a group under multiplication, $G = F - \{0\}$ is a group of order $q - 1$ under multiplication. For each $a \in G$, $a^{q-1} = 1 \Rightarrow a^q = a$.

Trivially $0^q = 0$. Thus for all $b \in F$, $b^q = b$. As a consequence every element of F is a root of $x^q - x$. If P is the prime subfield of F , then $x^q - x \in P[x]$ cannot have more than q roots in any extension of P (Theorem 13.28). So all the roots of $x^q - x$ are in F . Hence F is the splitting field of $x^q - x$ over P . ■

Theorem 13.54. (E.H. Moore). *For every prime number p and $n \in \mathbb{N}$, there exists a field having p^n elements.*

Proof: Consider $f(x) = x^{p^n} - x$. Clearly $f(x) \in \mathbb{Z}_p[x]$. Let K be the splitting field of $f(x)$ over $\mathbb{Z}_p$. We claim that K is a field having p^n elements. Now $x^{p^n} - x = \sum_{i=1}^{p^n} (x - a_i)$, $a_i \in K$. The mapping $\sigma : K \rightarrow K$

given by $\sigma(y) = y^{p^n} \forall y \in K$ is a monomorphism. (Cf. Lemma 13.46 and Problem 7). As a consequence $\sigma(K)$, ($\cong K$) is a subfield of K .

Now for each i ($1 \leq i \leq p^n$), $a_i^{p^n} = a_i \Rightarrow a_i = \sigma(a_i) \Rightarrow a_i \in \sigma(K)$. Thus every root of $f(x)$ is in $\sigma(K)$; also $\sigma(K)$ contains the prime subfield $\mathbb{Z}_p$ of K (Prove!). So $\sigma(K)$ contains the splitting field of $f(x)$ over $\mathbb{Z}_p$. This forces $\sigma(K) = K$. Note that each element of $\sigma(K)$ is a root of $f(x)$, so $K = \{a_1, a_2, \dots, a_{p^n}\}$. We assert that all a_i 's are distinct. Now $f'(x) = p^n x^{p^n-1} - 1 = -1$ as $\text{Char } K = p$. So $f(x)$ has no multiple roots (Lemma 13.41). Thus our assertion is proved.

Consequently K has exactly p^n elements. ■

Next theorem shows that for every prime number p and $n \in \mathbb{N}$, there is a unique field (within isomorphism) with p^n elements.

Theorem 13.55. *Finite fields having the same number of elements are isomorphic.*

Proof: Let K_1 and K_2 be two fields each having q elements. By Theorem 13.52, $q = p^n$ for some $n \in \mathbb{N}$ and prime number p , further characteristic of each of K_1 and K_2 is p . Let P_1 and P_2 be the prime subfields of K_1 and K_2 respectively. As $P_1 \cong \mathbb{Z}/\langle p \rangle$ and $P_2 \cong \mathbb{Z}/\langle p \rangle$, $P_1 \cong P_2$. Now by Lemma 13.53, K_1 is splitting field of $x^q - x$ over P_1 and K_2 is splitting field of $y^q - y$ over P_2 . Hence by Theorem 13.35, $K_1 \cong K_2$. ■

NOTATION. A finite field of q elements is denoted by $GF(q)$. Lastly we prove an important characterization of finite fields.

Theorem 13.56. *A field is finite if and only if its multiplicative group is cyclic.*

Proof: Let F be a field with q elements. Then $G = F - \{0\}$ is an Abelian group of order $q-1$. Let n be the exponent of G . By definition of exponent, $n \leq q-1$. Also $x^n = 1$ for all $x \in G$. But $x^n - 1$ can have at most n roots in G (Theorem 13.28), so $q-1 \leq n$. Hence $n = q-1$. Consequently there exists an element $a \in G$ such that $o(a) = q-1$. In other words G is cyclic.

Suppose $G = F - \{0\}$ is a cyclic group generated by a . If $a = 1$, we are done, since then $o(G) = 1 \Rightarrow o(F) = 2$. Let us now discuss the case $a \neq 1$.

Suppose that $\text{char } F = 0$. $-1 \in G$ and $-1 \neq 1 \Rightarrow -1 = a^n$ for some $n (\neq 0) \in \mathbb{Z} \Rightarrow (-1)^2 = a^{2n} = a^{2n} = 1$ with $n \neq 0 \Rightarrow G$ is a finite group. This in turn implies that F is finite, which is absurd as a field of characteristic zero cannot be finite. Hence $\text{char } F \neq 0$. So we have $\text{char } F = p$, where p is a prime number. The prime subfield P of F , being isomorphic to $\mathbb{Z}_p$, has p elements. Since $a - 1 \neq 0 \Rightarrow a - 1 \in G$, we get that $a - 1 = a^n$ for some $n \in \mathbb{Z}$. Hence a satisfies some polynomial in $P[x]$. In other words a is algebraic over P . Consequently $[P(a) : P] = r$ for some $r \in \mathbb{N}$. As P has p elements, $P(a)$ has p^r elements. Now $0 \in P(a)$ and every non-zero element of F , being a power of a also belongs to $P(a)$. Thus $F \subseteq P(a)$. But $P(a) \subseteq F$. Hence $F = P(a)$ has p^r elements. ■

WORKED-OUT EXERCISES

Exercise 1. A field F is called **perfect field** if all finite extensions of F are separable.

Prove that an algebraic extension of a perfect field is separable.

Solution. Let $a \in K$, an algebraic extension of F . Since a is algebraic over F , $[F(a) : F] = n$ for some $n \in \mathbb{N}$. This implies that $F(a)$ is a finite extension of F . By hypothesis F is perfect, so $F(a)$ is separable extension of F . As a consequence a is separable over F . Hence K is a separable extension of F .

Exercise 2. Let F be a field of characteristic $p (\neq 0)$. Show that an element a , in some extension of F , is separable over F if and only if $F(a^p) = F(a)$.

Solution. Let K be an extension of F and $a \in K$ be separable over F . The minimal polynomial $f(x) = x^n + \alpha_1 x^{n-1} + \dots + \alpha_{n-1} x + \alpha_n$ of a over F is separable. Let $g(x) = x^{np} + \alpha_1^p x^{n(p-1)} + \dots + \alpha_{n-1}^p x^{p(n-1)} + \alpha_n^p$. Then a is a root of $g(x)$ as $g(a^p) = \alpha_0^p + \alpha_1^p a^{p^2} + \dots + \alpha_{n-1}^p a^{p(n-1)} + \alpha_n^p = (\alpha_0 + \alpha_1 a + \dots + \alpha_{n-1} a^{n-1} + \alpha_n)^p = [f(a)]^p = 0$. Also $g(x)$ is irreducible over F . In fact if $h(x)$ is a factor of $g(x)$ in $F[x]$ then $h(x^p)$ is a factor of $g(x^p)$ in $F[x]$. But $g(x^p) = [f(x)]^p$ and $f(x)$ is irreducible over F .

$\Rightarrow h(x^p) = [f(x)]^k$ for some $0 \leq k \leq p$. Taking derivatives of both sides we get $k[f(x)]^{k-1} f'(x) = 0$. Separability of $f(x)$ forces $k=0$ or $k=p$. Accordingly $h(x^p) = 1$ or $h(x^p) = [f(x)]^p$. In the former case $h(x) = 1$ and in the latter case $h(x^p) = g(x^p) \Rightarrow h(x) = g(x)$. Thus $g(x)$ is the minimal polynomial of a^p over F and so $[F(a^p) : F] = n = [F(a) : F]$. Consequently $F(a^p) = F(a)$.

Conversely, let $F(a^p) = F(a)$ and suppose that a is not separable over F . The minimal polynomial $f(x)$ of a over F is not separable. This gives that $f(x) = g(x^p)$ and so a^p is a root of $g(x)$. Clearly degree

of $g(x)$ is $\frac{n}{p} = m$ (say). Hence $[F(a^p) : F] \leq m < n$, as $p > 1$. Since $[F(a) : F] = [F(a) : F(a^p)][F(a^p) : F] \leq m$ (by hypothesis), we get that $n \leq m$. This is absurd. Hence a is separable.

Exercise 3. If $a \in K$, an extension of F , is separable over a field F of characteristic $p (\neq 0)$, then $F(a)$ is a separable extension of F .

Solution. Let $b \in F(a)$. As a is algebraic over F , $[F(a) : F]$ is finite. Let $[F(b) : F] = m$, $[F(a) : F(b)] = n$, $[F(a) : F(b^p)] = q$ and $[F(b^p) : F] = s$. Now $nm = [F(a) : F(b)][F(b) : F] = [F(a) : F] = [F(a) : F(b^p)][F(b^p) : F] = qs$. Again $F(b^p)$ is a subfield of $F(b)$ implies that $s \leq m \Rightarrow n \leq q$.

Let $f(x) = \alpha_0 + \alpha_1 x + \dots + x^n$ be the minimal polynomial of a over $F(b)$. As $\{1, b, \dots, b^{n-1}\}$ is a basis of $F(b)$ over F ; for each i , $0 \leq i \leq n-1$, $\alpha_i = \lambda_{i,0} + \lambda_{i,1}b + \dots + \lambda_{i,m-1}b^{m-1}$ with $\lambda_{i,j} \in F$ for all $0 \leq j \leq m-1$. Further $f(a) = 0 \Rightarrow a^p$ is a root of $g(x) = \alpha_0^p + \alpha_1^p x + \dots + x^n$. $0 \leq j \leq m-1$. For each i , $0 \leq i \leq n-1$; $\alpha_i^p = (\lambda_{i,0} + \lambda_{i,1}b + \dots + \lambda_{i,m-1}b^{m-1})^p = \lambda_{i,0}^p + \lambda_{i,1}^p b^p + \dots + \lambda_{i,m-1}^p b^{(m-1)p} \in F(b^p)$. Hence $g(x) \in F(b^p)[x]$. But a is separable is satisfied by a^p . As a result $[F(a^p, b^p) : F(b^p)] \leq n$. But a is separable over $F \Rightarrow F(a^p) = F(a)$ (Exercise 2). So $[F(a) : F(b^p)] \leq n$, i.e., $q \leq n$. [Note that $F(a^p) = F(a)$ and $b^p \in F(a)$ imply that $F(a^p, b^p) = F(a^p)$.]

Consequently $q = n$ and $F(b) = F(b^p) \Rightarrow b$ is separable over F . Hence $F(a)$ is separable extension of F .

Exercise 4. $a, b \in K$, an extension of F , are separable over F . Prove that $F(a, b)$ is separable extension of F .

Solution. In case F is finite, $F(a, b)$ is separable extension by Theorem 13.40. In case $\text{char } F = 0$, there exists $c \in F(a, b)$ such that $F(a, b) = F(c)$. $\text{Char } F = 0$ implies that irreducible polynomial of c over F is separable. Hence c is separable, *a fortiori* $F(a, b)$ is separable extension of F .

So the only case left is that of infinite field of characteristic $p (\neq 0)$. Following the arguments given in the proof of Theorem 13.49, one can easily see that $F(a, b)$ is simple extension of F , i.e., $F(a, b) = F(c)$ where $c = a + \lambda b$, $\lambda \in F$.

Now let $[F(c) : F] = m$. We can write $a = \alpha_0 + \alpha_1 c + \dots + \alpha_{m-1} c^{m-1}$ and $b = \beta_0 + \beta_1 c + \dots + \beta_{m-1} c^{m-1}$ with $\alpha_i, \beta_i \in F$ for $i = 0, 1, \dots, m-1$. Clearly $a^n = \alpha_0^n + \alpha_1^n c^n + \dots + \alpha_{m-1}^n c^{(m-1)n} \in F(c^n)$ and similarly $b^n \in F(c^n)$. As a, b are separable over F , we get that $F(a) = F(a^n) \subseteq F(c^n)$ and $F(b) = F(b^n) \subseteq F(c^n) \Rightarrow a, b \in F(c^n)$. This in turn implies that $c = a + \lambda b \in F(c^n)$, i.e., $F(c) \subseteq F(c^n)$. Trivially $F(c^n) \subseteq F(c)$. So $F(c) = F(c^n) \Rightarrow c$ is separable over F . Invoking the result of Exercise 3, we get that $F(a, b)$ is separable extension of F .

Exercise 5. (Steinitz Theorem). If K is a finite extension of F then K is simple extension of F if and only if there are only a finite number of subfields of K containing F .

Solution. Suppose that K is simple extension of F . Then $K = F(c)$ for some $c \in K$. Let $f(x)$ be the minimal polynomial of c over F and L , a subfield of F , containing K . As c is algebraic over L , let $g(x) = \alpha_0 + \alpha_1 x + \dots + \alpha_{n-1} x^{n-1} + x^n$ be the minimal polynomial of c over L . As $f(x) \in F[x] \Rightarrow f(x) \in L[x]$, we get that $g(x)$ divides $f(x)$ in $L[x]$, so in $K[x]$. Put $L_0 = F(\alpha_0, \dots, \alpha_{n-1})$. Thus $g(x) \in L_0[x]$; further $L_0 \subseteq L$ gives that g is irreducible over L_0 . Hence $g(x)$ is minimal polynomial of c over L_0 . Now note that $K = F(c) \subseteq L(c) \Rightarrow K = L(c)$ and so $[K : L] = n$. Similar arguments show that $K = L_0(c)$ and as minimal polynomial of c over L_0 is $g(x)$ we get that $[K : L_0] = n$. Hence $L = L_0$. Thus the only subfields of K containing F are the fields of type $F(\beta_0, \beta_1, \dots, \beta_{n-1})$ where $\beta_0 + \beta_1 x + \dots + \beta_{n-1} x^{n-1} + x^n$ is a factor of $f(x)$ in $K[x]$. As number of monic factors of $f(x)$ in $K[x]$ is finite we get that the number of subfields of K containing F is finite.

Suppose now the number of subfields of K containing F is finite. In case F is finite; K being a finite extension of F , is also finite. Hence $K - \{0\}$ is a cycle group under multiplication. Let $K - \{0\} = \langle c \rangle$. Then $K = F(c)$ is simple extension of F . So let us now assume that F is an infinite field. Since for each $a \in K$, $F(a)$ is a subfield of K containing F and number of subfields of K containing F is finite, we can choose $b \in F$, such that $[F(b) : F] = n$ and n is the largest among all integers $[F(a) : F]$, $a \in K$. We claim that $K = F(b)$. If our claim is not true, we can find $c \in K$ such that $c \notin F(b)$. By hypothesis the set of fields $F(cd + b)$ where $d \in F$, is finite. As F is infinite, there exist $r, s \in F$, $r \neq s$ such that $F(cr + b) = F(cs + b)$. Let $z = cr + b$. Then $F(z) = F(cs + b) \Rightarrow cr + b, cs + b \in F(z) \Rightarrow c(r - s) = (cr + b) - (cs + b) \in F(z) \Rightarrow c \in F(z)$ as $r - s \neq 0$. Further $b = (cr + b) - cr \in F(z) \Rightarrow F(b) \subseteq F(z)$. But $c \in F(z)$ and $c \notin F(b)$; consequently $F(b) \neq F(z)$, i.e., $[F(z) : F(b)] > 1$. This in turn gives that $[F(z) : F] = [F(z) : F(b)][F(b) : F] > n$, a contradiction to the choice of n .

Hence $K = F(b)$ is a simple extension of F .

PROBLEMS

1. Write in details the proof of Lemma 13.39.
2. Let F be a field of characteristic $p > 0$, a prime number. Consider $f(x) = x^p - x$. Show that

(i) $f'(x) = -1$.

(ii) $f(x)$ and $f'(x)$ have no common roots.

(iii) All roots of $f(x)$ in its splitting field over F are distinct.

[Hint. For (ii) notice that $f'(x)$ has no roots, and (iii) follows from (ii)].

3. If K is a field extension of $\mathbb{Z}_p$ then shows that $\mathbb{Z}_p$ is the prime subfield of K .

4. Prove that any finite extension of a finite field is simple. (Another special case of the theorem of primitive elements).

5. Let $f(x) \in F[x]$ be an irreducible polynomial. Prove that the HCF of $f(x)$ and $f'(x)$ is 1 if and only if $f'(x) \neq 0$.

[Hint. Use Lemma 13.41].

6. Prove that there is no rational function $\frac{f(t)}{g(t)}$, $f(t), g(t) \in F[t]$, $g(t) \neq 0$ such that, $t = \left[\frac{f(t)}{g(t)} \right]^2$, where t is an indeterminate over F .

Deduce that $x^2 - t$ is irreducible over the field $\mathbb{Z}_2(t)$.

7. Let F be a field of characteristic p , then prove that for all $a, b \in F$ and for all non-negative integers n ,

(i) $(a+b)^{p^n} = a^{p^n} + b^{p^n}$

(ii) $(a-b)^{p^n} = a^{p^n} - b^{p^n}$.

[Hint. Use induction on n].

8. Let F be a finite field and n be any positive integer. Prove that there exists at least one irreducible polynomial of degree n over F .

[Hint. Let K be the splitting field of $x^{q^n} - x$ over F where q is the order of field F . Then $[K : F] = n$, gives $K = F(c)$ for some $c \in K$. Then the minimal polynomial of c over F is a desired polynomial].

9. Prove that any field which is either of characteristic zero or is a finite field is perfect.

10. For any three fields $F \in L \subset K$ prove that if K is separable over F then K is separable over L .

11. If K is an extension of F prove that the set of elements in K which are separable over F form a subfield of K .

12. Find a primitive element in each of the following fields over $\mathbb{Q}$.

(i) $\mathbb{Q}(\sqrt{2}, \sqrt{3})$.

(ii) $\mathbb{Q}(i, \sqrt{2})$.

(iii) $\mathbb{Q}(\sqrt{2}, \sqrt[3]{5})$.

(iv) $\mathbb{Q}(\sqrt{-3}, \sqrt{2})$.

(v) $\mathbb{Q}(\omega, \sqrt{2})$ where ω is imaginary cube root of unity.

(vi) $\mathbb{Q}(\omega, \sqrt[3]{2})$.

13. Let F be a field of characteristic $p > 0$ and $a \in F$ be such that there exists no $b \in F$ with $b^p = a$ (i.e. $\sqrt[p]{a} \notin F$). Then show that $x^p - a$ is irreducible and inseparable over F .

14. Deduce from problem 13 that a field F of characteristic $p (\neq 0)$ is perfect if and only if the mapping $a \rightarrow a^p$, $a \in F$ is an automorphism of F .

15. Deduce from problem 13 that for any field F of characteristic $p (\neq 0)$ if π is transcendental over F , the $F(\pi)$ is not perfect.

16. Let F be of characteristic $p > 0$ and $f(x)$ be a non-constant polynomial over F . Prove the following.

(i) There exist integer $n \geq 0$ and $g(x) \in F[x]$ such that $f(x) = g(x^{p^n})$ and $g'(x) \neq 0$.

(ii) If $f(x)$ is irreducible then $g(x)$ is irreducible.

(iii) If $f(x)$ is irreducible then $g(x)$ is separable.

(iv) If $f(x)$ is irreducible then each root of $f(x)$ has multiplicity p^n in the splitting field of $f(x)$.

(v) If an element a in a field extension K of F is algebraic over F then there exists a non-negative integer n such that a^{p^n} is separable over F .

17. If f is a function from a field F to F such that $f(x) = x^{-1}$ if $x \neq 0$ and $f(0) = 0$, then show that f is an automorphism if and only if F has either 2, 3, 4 elements.

18. Let F be a finite field of characteristic different from 2. Show that a non-zero element a of F is square of some element in F if and

only if $a^{\frac{q-1}{2}} = 1$ and a non-zero element a of F is not a square of

any element in F if and only if $a^{\frac{q-1}{2}} = -1$ where $o(F) = q$.

19. Let $\mathbb{Z}_2$ be the ring of integers modulo 2, and x, y , transcendental over $\mathbb{Z}_2$. Consider $\mathbb{Z}_2(x, y)$, the field of quotients of $\mathbb{Z}_2[x, y]$. Take F to be subfield $\mathbb{Z}_2(x^2, y^2)$ of $K = \mathbb{Z}_2(x, y)$. Show that (i) $\{1, x, y, xy\}$ is a basis of K over F , and (ii) every element of K satisfies a polynomial over F of degree at most 2. Deduce that K is not a simple extension of F .

20. Prove that a finite separable extension of every field F is a simple extension.

21. Let $a, b \in K$, an extension of F , be algebraic over F . If one of a or b is separable over F , prove that $F(a, b)$ is a simple extension of F .

GALOIS THEORY

One of the most elegant theories in Abstract Algebra is the "Galois theory of fields". This theory is an excellent composite of the theory of groups and the theory of algebraic field extensions. It has many applications to the Theory of Equations and Geometry. In an introductory text on Abstract Algebra, like the present one, it is not possible to discuss in details the whole of Galois theory. The interested readers can find its excellent accounts in some of the books in the bibliography. However we shall study its fundamental concepts and give some simple applications.

1. Monomorphisms and their Linear Independence

Let S be any set and K be any field. Let $M(S, K)$ denote the set of all mappings of S into K . Given $\phi_1, \phi_2 \in M(S, K)$ and $a \in K$, we define

$$(\varphi_1 + \varphi_2)s = \varphi_1(s) + \varphi_2(s)$$

$$(a\varphi_1)s = a\varphi_1(s) \text{ for every } s \in S.$$

Then $\varphi_1 + \varphi_2 \in M(S, K)$ and $a\varphi_1 \in M(S, K)$. Under these operations of addition and multiplication, $M(S, K)$ becomes a vector space over K . The zero vector in $M(S, K)$ is the mapping $\bar{0} : S \rightarrow K$ given by $\bar{0}(s) = 0$ for all $s \in S$. So we can say that $\phi_1, \phi_2, \dots, \phi_n \in M(S, K)$ are linearly independent over K if for any $a_1, a_2, \dots, a_n \in K$

$$a_1\phi_1 + a_2\phi_2 + \dots + a_n\phi_n = \bar{0}$$

implies $a_i = 0 \forall i = 1, 2, \dots, n$, i.e. $a_1\varphi_1(s) + a_2\varphi_2(s) + \dots + a_n\varphi_n(s) = 0$ for all $s \in S$ implies $a_i = 0 \forall i = 1, 2, \dots, n$.

In particular consider a field E and the vector space $M(E, K)$. Let $\sigma_1, \sigma_2, \dots, \sigma_n$ be any n distinct monomorphisms of E into K . Clearly $\sigma_i \in M(E, K) \forall i = 1, 2, \dots, n$. So we can talk of their linear independence or dependence over K .

Theorem 14.1. *Let E and K be any two fields. If $\sigma_1, \sigma_2, \dots, \sigma_n$ are any n distinct monomorphisms of E into K , then they are linearly independent over K .*

Proof: We prove the result by induction on n . Now for any $a_1 \in K$, $a_1 \sigma_1 = \bar{0} \Rightarrow a_1 \sigma_1(x) = 0$ for all $x \in E \Rightarrow a_1 = 0$ as $\sigma_1(x) \neq 0$ for all $x (\neq 0) \in E$. Thus the result holds for $n=1$. Let us suppose that the result holds for any set of monomorphisms of E into K having less than n members. Suppose for some $a_i \in K$, $i=1, 2, \dots, n$ not all zero,

$$a_1 \sigma_1 + a_2 \sigma_2 + \dots + a_n \sigma_n = \bar{0} \quad \dots (1)$$

The induction hypothesis implies none of the a_i 's is zero.

As $a_n \neq 0$; dividing (1) by a_n , we get

$$b_1 \sigma_1 + b_2 \sigma_2 + \dots + b_{n-1} \sigma_{n-1} + \sigma_n = \bar{0} \quad \dots (2)$$

where

$$b_i = a_i a_n^{-1} \quad \forall i=1, 2, 3, \dots, n-1.$$

Since σ_1 and σ_n are distinct, there exists $x_1 \in E$ such that $\sigma_1(x_1) \neq \sigma_n(x_1)$. Clearly $x_1 \neq 0$. For any $x \in E$, we have from (2)

$$b_1 \sigma_1(x_1 x) + b_2 \sigma_2(x_1 x) + \dots + \sigma_n(x_1 x) = 0$$

$$i.e. \quad b_1 \sigma_1(x_1) \sigma_1(x) + b_2 \sigma_2(x_1) \sigma_2(x) + \dots + \sigma_n(x_1) \sigma_n(x) = 0 \quad \dots (3)$$

$$\text{Thus } b_1 \frac{\sigma_1(x_1)}{\sigma_n(x_1)} \sigma_1(x) + b_2 \frac{\sigma_2(x_1)}{\sigma_n(x_1)} \sigma_2(x) + \dots + \sigma_n(x) = 0 \quad \forall x \in E \quad \dots (4)$$

$$(2) \text{ gives } b_1 \sigma_1(x) + b_2 \sigma_2(x) + \dots + \sigma_n(x) = 0 \quad \forall x \in E. \quad \dots (5)$$

Subtracting (5) from (4), we get

$$b_1 \left[\frac{\sigma_1(x_1)}{\sigma_n(x_1)} - 1 \right] \sigma_1(x) + b_2 \left[\frac{\sigma_2(x_1)}{\sigma_n(x_1)} - 1 \right] \sigma_2(x) + \dots$$

$$\dots + b_{n-1} \left[\frac{\sigma_{n-1}(x_1)}{\sigma_n(x_1)} - 1 \right] \sigma_{n-1}(x) = 0 \quad \forall x \in E.$$

Since $b_1 \left[\frac{\sigma_1(x_1)}{\sigma_n(x_1)} - 1 \right] \neq 0$. The last equation implies that $\sigma_1, \sigma_2, \dots, \sigma_{n-1}$ are linearly dependent over K . This contradicts the induction hypothesis.

Hence $\sigma_1, \sigma_2, \dots, \sigma_n$ are linearly independent over K . This proves the theorem. ■

Taking $E=K$, we get the following:

Corollary 14.2. Any set of automorphisms of K is linearly independent over K . ■

Lemma 14.3. The set of all automorphisms of a field form a group under resultant composition.

Proof: Let $\text{Aut}(K)$ be the set of all automorphisms of K . Since I , the identity map on K , is an automorphism of K ; $I \in \text{Aut}(K)$.

(i) *Closure:* Let $\sigma_1, \sigma_2 \in \text{Aut}(K)$. Then σ_1 and σ_2 are both 1-1 and onto mappings. So $\sigma_1 \sigma_2$ is also 1-1 and onto mapping. Further for any $x, y \in K$

$$\begin{aligned} \sigma_1 \sigma_2(x+y) &= \sigma_1[\sigma_2(x+y)] = \sigma_1[\sigma_2(x) + \sigma_2(y)] \\ &= \sigma_1[\sigma_2(x)] + \sigma_1[\sigma_2(y)] = \sigma_1 \sigma_2(x) + \sigma_1 \sigma_2(y) \\ &= \sigma_1[\sigma_2(x)] + \sigma_1[\sigma_2(y)] = \{\sigma_1[\sigma_2(x)]\} + \{\sigma_1[\sigma_2(y)]\} \\ \sigma_1 \sigma_2(xy) &= \sigma_1[\sigma_2(xy)] = \sigma_1[\sigma_2(x) \sigma_2(y)] = \{\sigma_1[\sigma_2(x)]\} \{\sigma_1[\sigma_2(y)]\} \\ &= [\sigma_1 \sigma_2(x)] [\sigma_1 \sigma_2(y)] \end{aligned}$$

This shows that $\sigma_1\sigma_2$ is an automorphism of K .

Hence $\sigma_1\sigma_2 \in \text{Aut}(K)$.

(ii) *Associativity* : It follows from the fact that the resultant composition of mappings, in general, is associative.

(iii) *Existence of Identity* : The identity map I on K is the identity of $\text{Aut}(K)$.

(iv) *Existence of Inverse* : Consider any $\sigma \in \text{Aut}(K)$. Since σ is 1-1 and onto, σ^{-1} (the inverse of the mapping σ) exists. For any $x \in G$, $\sigma^{-1}(x) = y$ if and only if $\sigma(y) = x$. Now consider $x_1, x_2 \in K$. Let $\sigma^{-1}(x_1) = y_1$, $\sigma^{-1}(x_2) = y_2$. Then $\sigma(y_1) = x_1$, $\sigma(y_2) = x_2$ so that $\sigma(y_1 + y_2) = \sigma(y_1) + \sigma(y_2) = x_1 + x_2$ and $\sigma(y_1 y_2) = \sigma(y_1)\sigma(y_2) = x_1 x_2$. Hence $\sigma^{-1}(x_1 + x_2) = y_1 + y_2 = \sigma^{-1}(x_1) + \sigma^{-1}(x_2)$ and $\sigma^{-1}(x_1 x_2) = y_1 y_2 = \sigma^{-1}(x_1)\sigma^{-1}(x_2)$. Consequently σ^{-1} is an automorphism of K i.e. $\sigma^{-1} \in \text{Aut}(K)$.

This proves that $\text{Aut}(K)$ is a group. ■

Definition 14.4. Let F be any field and K be any field extension of F . Then an automorphism σ of K is said to be an F -automorphism if $\sigma(x) = x$ for every $x \in F$ (i.e. σ leaves every element of F fixed).

Clearly I , the identity automorphism of K is an F -automorphism.

Let σ_1, σ_2 be any two F -automorphisms of K . Then $\sigma_1\sigma_2^{-1} \in \text{Aut}(K)$ as $\text{Aut}(K)$ is a group. At the same time for any $x \in K$, $\sigma_1(x) = x$, $\sigma_2(x) = x$. This implies $\sigma_1(x) = x$, $\sigma_2^{-1}(x) = x \Rightarrow \sigma_1\sigma_2^{-1}(x) = \sigma_1[\sigma_2^{-1}(x)] = \sigma_1(x) = x$. Hence $\sigma_1\sigma_2^{-1}$ is also an F -automorphism of K . Thus the set of all F -automorphisms of K is a subgroup of all automorphisms of K .

NOTATION. $G(K, F)$ will denote the group of all F -automorphisms of K . $G(K, F)$ is called the *Galois group of K over F* .

Lemma 14.5. Let K be any field extension of F and $a \in K$ be algebraic over F . Then for every F -automorphism σ of K , $\sigma(a)$ is a conjugate of a over F .

Proof : Let $f(x) = x^n + \alpha_{n-1}x^{n-1} + \alpha_{n-2}x^{n-2} + \dots + \alpha_0$ be the minimal polynomial of a over F .

Then $a^n + \alpha_{n-1}a^{n-1} + \alpha_{n-2}a^{n-2} + \dots + \alpha_0 = 0$

So $0 = \sigma(0) = \sigma(a^n + \alpha_{n-1}a^{n-1} + \alpha_{n-2}a^{n-2} + \dots + \alpha_0)$
 $= [\sigma(a)]^n + \sigma(\alpha_{n-1})[\sigma(a)]^{n-1} + \sigma(\alpha_{n-2})[\sigma(a)]^{n-2} + \dots + \sigma(\alpha_0)$
 $= [\sigma(a)]^n + \alpha_{n-1}[\sigma(a)]^{n-1} + \alpha_{n-2}[\sigma(a)]^{n-2} + \dots + \alpha_0,$

since $\sigma(\alpha_i) = \alpha_i \forall i = 0, 1, 2, \dots, n-1$.

This shows that $\sigma(a)$ is also a root of $f(x)$. Hence $\sigma(a)$ is a conjugate of a over F . ■

REMARK : Let $K = F(a_1, a_2, \dots, a_n)$ be a finite algebraic extension with $\{a_1, a_2, \dots, a_n\}$ as a basis of K . Then each $x \in K$ is expressible

as $x = \alpha_1 a_1 + \alpha_2 a_2 + \dots + \alpha_n a_n$.

Suppose σ is any F -automorphism of K . Then

$$\begin{aligned}\sigma(x) &= \sigma(\alpha_1)\sigma(a_1) + \sigma(\alpha_2)\sigma(a_2) + \dots + \sigma(\alpha_n)\sigma(a_n) \\ &= \alpha_1\sigma(a_1) + \alpha_2\sigma(a_2) + \dots + \alpha_n\sigma(a_n).\end{aligned}$$

So that $\sigma(x)$ is known if we know $\sigma(a_1), \sigma(a_2), \dots, \sigma(a_n)$ i.e. σ is completely determined by the images of the basis elements of K . In fact, more generally, if K is finitely generated over F and $a_1, a_2, \dots, a_n$ is a set of generators of K over F then σ is determined by $\sigma(a_1), \sigma(a_2), \dots, \sigma(a_n)$.

Lemma 14.6. Let G be a group of automorphisms of a field K . Then the set $F_0 = \{x \in K \mid \sigma(x) = x \forall \sigma \in G\}$ is a subfield of K .

Proof: Since $\sigma(0) = 0$ and $\sigma(1) = 1 \forall \sigma \in G$; $0, 1 \in F_0$.

Let $x, y \in F_0$. Then for every $\sigma \in G$, $\sigma(x) = x$, $\sigma(y) = y$. Consequently $\sigma(x - y) = \sigma(x) - \sigma(y) = x - y$ and when $y \neq 0$, $\sigma(xy^{-1}) = \sigma(x)\sigma(y)^{-1} = x\sigma(y)^{-1} = xy^{-1}$. Thus $x - y \in F_0$ and when $y \neq 0$, $xy^{-1} \in F_0$. Hence F_0 is a subfield of K . ■

Definition 14.7. Let G be a group of automorphisms of K . The subfield F_0 of K consisting of all those elements $x \in K$ such that $\sigma(x) = x$ for every $\sigma \in G$, is called the *fixed field* under G . $\Gamma(G)$ or K_G will denote the fixed field under G .

REMARK 1. Let P be the prime subfield of K . We know that P is contained in every subfield of K . In particular if F_0 is the fixed field under G , then $P \subseteq F_0$. So that $\sigma(x) = x \forall x \in P$ and $\forall \sigma \in G$. Thus every automorphism of K is a P -automorphism.

REMARK 2. Every member of G can be regarded as an F_0 automorphism of K where $F_0 = \Gamma(G)$.

REMARK 3. If K is a field extension of a field F and G is a group of F -automorphisms of K then $F \subseteq F_0$, the fixed field under G .

Before we proceed with the theory, we pause to give some examples of the concepts introduced above.

EXAMPLE 1. Let $K = \mathbb{Q}(\sqrt{2})$. The minimal polynomial of $\sqrt{2}$ over $\mathbb{Q}$ is $x^2 - 2$ and $\{1, \sqrt{2}\}$ forms a basis of K over $\mathbb{Q}$. Thus any $x \in K$ is of the form $x = a + b\sqrt{2}$; $a, b \in \mathbb{Q}$. Let σ be any automorphism of K , $\sigma(a) = a \forall a \in \mathbb{Q}$. Further $\sigma(\sqrt{2})$ is a conjugate of $\sqrt{2}$ over $\mathbb{Q}$. As $\pm\sqrt{2}$ are the only roots of $x^2 - 2$ we get $\sigma(\sqrt{2}) = \sqrt{2}$ or $\sigma(\sqrt{2}) = -\sqrt{2}$.

Case I: $\sigma(\sqrt{2}) = \sqrt{2}$. Then $\sigma(x) = \sigma(a) + \sigma(b)\sigma(\sqrt{2}) = a + b\sqrt{2} = x \forall x \in K$, gives $\sigma = I$, the identity map on K .

Case II: If $\sigma(\sqrt{2}) = -\sqrt{2}$, then $\sigma(x) = a - b\sqrt{2}$. Hence $\text{Aut}(K)$ consists of only two elements namely I and the

automorphism η such that $\eta(x) = a - b\sqrt[3]{2} \forall x = a + b\sqrt[3]{2} \in K$.

Suppose x is in the fixed field F_0 under $\text{Aut}(K)$. Then $\eta(x) = x$ i.e. $a - b\sqrt[3]{2} = a + b\sqrt[3]{2}$. Thus $b = 0$ and $x = a \in \mathbb{Q}$. Hence $F_0 \subseteq \mathbb{Q}$. However $\mathbb{Q} \subseteq F_0$ as $\mathbb{Q}$ is a prime field. Hence $F_0 = \mathbb{Q}$. So $\mathbb{Q}$ is the fixed field under $\text{Aut}(K)$.

EXAMPLE 3. Let $K = \mathbb{Q}(\sqrt[3]{2})$. The minimal polynomial of $\sqrt[3]{2}$ over $\mathbb{Q}$ is $x^3 - 2$. It has only one real root namely $\sqrt[3]{2}$. K is a field of real numbers. So $\sqrt[3]{2}$ is the only root of $x^3 - 2$ in K . Let σ be any $\mathbb{Q}$ automorphism of K . Then $\sigma(\sqrt[3]{2}) \in K$ is a root of $x^3 - 2$. So $\sqrt[3]{2} = \sigma(\sqrt[3]{2})$. Let x be any element of K . x can be expressed as $\alpha + (\sqrt[3]{2})\beta + (\sqrt[3]{2})^2\gamma$; $\alpha, \beta, \gamma \in \mathbb{Q}$.

$$\begin{aligned}\text{So } \sigma(x) &= \sigma(\alpha) + \sigma(\sqrt[3]{2})\sigma(\beta) + [\sigma(\sqrt[3]{2})]^2\sigma(\gamma) \\ &= \alpha + \sqrt[3]{2}\beta + (\sqrt[3]{2})^2\gamma = x.\end{aligned}$$

That means $\sigma = I$. Thus $\text{Aut}(K) = \{I\}$.

Hence in this case K itself is the fixed field under $\text{Aut}(K)$.

EXAMPLE 3. Let L be any field of characteristic 2 and t be an indeterminate over L . Let $F = L(t)$ and K be the splitting field of $x^2 - t$ over F . If $a \in K$ is a root of $x^2 - t$, then $t = a^2$ and $x^2 - t = x^2 - a^2 = (x - a)^2$. This gives that a is the only conjugate of a over F . Thus for any $\sigma \in G(K, F)$, $\sigma(a) = a$ (Lemma 14.5). Take $x \in K$ then $x = \alpha + a\beta$ for some $\alpha, \beta \in F$ as $K = F(a)$. So $\sigma(x) = \sigma(\alpha) + \sigma(a)\sigma(\beta) = \alpha + a\beta = x$. Hence $\sigma = I$ and $\text{Aut } K = \{I\}$. In this case the fixed field under $G(K, F)$ is K itself.

Theorem 14.8. (Artin) Let G be a finite subgroup of automorphisms of a field K ; F_0 , the fixed field under G . Then the degree of K over F_0 is equal to the order of the group G .

Proof: Let $o(G) = n$ we show that

- (i) if $[K : F_0]$ is finite, say m , then $m \geq n$ and
- (ii) $[K : F_0]$ is finite, say m , such that $m \leq n$.
- (i) Suppose $m < n$. Let $\sigma_1 = I, \sigma_2, \dots, \sigma_n$ be all the members of G . Let $\{x_1, x_2, \dots, x_m\}$ be a basis of K over F_0 . Consider the system of m linear homogeneous equations

$$\sigma_1(x_j)u_1 + \sigma_2(x_j)u_2 + \dots + \sigma_n(x_j)u_n = 0 \quad \forall j = 1, 2, \dots, m \quad \dots (1)$$

As the number of equations is less than the number of variables, the equations (1) have a non-trivial solution, say, $(y_1, y_2, \dots, y_n)$

$$\text{Then } \sigma_1(x_j)y_1 + \sigma_2(x_j)y_2 + \dots + \sigma_n(x_j)y_n = 0 \quad \forall j = 1, 2, \dots, m \quad \dots (2)$$

Consider any $x \in K$. Then $x = \alpha_1 x_1 + \alpha_2 x_2 + \dots + \alpha_m x_m$ for some $\alpha_i \in F_0$ as $\{x_1, x_2, \dots, x_m\}$ is a basis of K over F_0 . In (2), multiplying j th equation by α_j , adding the resulting equations and using the fact that $\sigma_i(\alpha_j) = \alpha_j \forall i, j$, we get

$$y_1 \sigma_1(x) + y_2 \sigma_2(x) + \dots + y_n \sigma_n(x) = 0 \quad \forall x \in K.$$

So that $y_1\sigma_1 + y_2\sigma_2 + \dots + y_n\sigma_n = \bar{0}$ with at least one of $y_j \neq 0$. This is not possible (Theorem 14.1). Hence $m \leq n$ so $m \geq n$.

(ii) Suppose there exist $n+1$ linearly independent elements say $x_1, x_2, \dots, x_{n+1}$ of K over F_0 . Consider the system of n linear homogeneous equations is $n+1$ unknowns.

$$\sigma_j(x_1)u_1 + \sigma_j(x_2)u_2 + \dots + \sigma_j(x_{n+1})u_{n+1} = 0 \text{ for } j=1, 2, \dots, n \dots (3)$$

As the number of variables is greater than the number of equations. These homogeneous equations have a non-trivial solution. Let $(z_1, z_2, \dots, z_{n+1})$ be a non-trivial solution of equations (3) with smallest number, say r , of non-zero components. We can renumber them and suppose that $z_j = 0 \forall j \geq r+1$.

Then (3) gives

$$\sigma_j(x_1)z_{1j} + \sigma_j(x_2)z_{2j} + \dots + \sigma_j(x_r)z_{rj} = 0 \forall j=1, 2, \dots, n \dots (4)$$

Dividing the equations by z_r and setting $z_i' = z_i/z_r \forall i$, we get

$$\sigma_i(x_j)z_1' + \sigma_i(x_2)z_2' + \dots + \sigma_i(x_{r-1})z_{r-1}' + \sigma_i(x_r) = 0 \forall j \dots (5)$$

Now for $j=1, \sigma_j = I$, so $\sigma_1(x_i) = x_i \forall i$. We get from (4)

$$x_1z_1' + x_2z_2' + \dots + x_{r-1}z_{r-1}' + x_r = 0. \dots (6)$$

If all of $z_1', z_2', \dots, z_{r-1}'$ were in F_0 then (6) would give $x_1, x_2, \dots, x_r$ are linearly dependent over F_0 . This would not be possible. Hence at least one of these, say, z_1' is not in F_0 . Notice further that $r \neq 1$ otherwise we would have got $z_1' = 0$. As $z_1' \notin F_0$ there exists some $\sigma_i \in G$ such that $\sigma_i(z_1') \neq z_1'$. Applying σ_i to (5) we get

$$\sigma_i[\sigma_j(x_1)z_1'] + \sigma_i[\sigma_j(x_2)z_2'] + \dots + \sigma_i[\sigma_j(x_{r-1})z_{r-1}'] + \sigma_i\sigma_j(x_r) = 0 \forall j=1, 2, \dots, n.$$

$$\text{i.e., } \sigma_i\sigma_j(x_1)\sigma_i(z_1') + \sigma_i\sigma_j(x_2)\sigma_i(z_2') + \dots + \sigma_i\sigma_j(x_{r-1})\sigma_i(z_{r-1}') + \sigma_i\sigma_j(x_r) = 0 \forall j=1, 2, \dots, n.$$

However $\sigma_i G = G$ so that every $\sigma_k \in G$ is the form $\sigma_i\sigma_j$, we get

$$\sigma_j(x_1)\sigma_i(z_1') + \sigma_j(x_2)\sigma_i(z_2') + \dots + \sigma_j(x_{r-1})\sigma_i(z_{r-1}') + \sigma_j(x_r) = 0 \dots (7)$$

for all $j=1, 2, \dots, n$.

Subtracting (7) from (5) we get

$$\sigma_i(x_1)[(z_1' - \sigma_i(z_1'))] + \sigma_i(x_2)[z_2' - \sigma_i(z_2')] + \dots + \sigma_i(x_{r-1})[z_{r-1}' - \sigma_i(z_{r-1}')] = 0 \forall j=1, 2, \dots, n \dots (8)$$

$$= 0 \forall k=1, 2, \dots, r-1.$$

Now we put $t_k = z_k' - \sigma_i(z_k') \forall k=1, 2, \dots, r-1$

and $t_k = 0 \forall k=r, r+1, \dots, n+1$

Then (8) gives

$$\sigma_i(x_1)t_1 + \sigma_i(x_2)t_2 + \dots + \sigma_i(x_{r-1})t_{r-1} + \sigma_i(x_r)t_r + \dots + \sigma_i(x_{n+1})t_{n+1} = 0 \forall j=1, 2, \dots, n; \dots (9)$$

$$+ \sigma_i(x_r)t_r + \dots + \sigma_i(x_{n+1})t_{n+1} = 0 \text{ so that } (t_1, t_2, \dots, t_{r-1}, 0, 0, \dots, 0)$$

Further $t_1 \neq 0$ as $z_1' \neq \sigma_i(z_1')$ so that $(t_1, t_2, \dots, t_{r-1}, 0, 0, \dots, 0)$ is a non-trivial solution of the system of equations (3). It has less than r non-zero components. But by our choice $(z_1, z_2, \dots, z_r, 0, 0, \dots, 0)$

is a non-trivial solution of (3), having smallest number of non-zero components and these are r in number. Hence we get a contradiction. This proves that $m \leq n$. Hence $m = n$, i.e., $[K : F_0] = o(G)$. This completes the proof. ■

REMARK 1. Let G be any group (not necessarily finite) of automorphisms of a field K such that $[K : F_0] = m$ is finite, where F_0 is the fixed field under G . In the first part of the proof of the above theorem, we don't require to assume that G is finite, actually we get that G cannot have more than m members.

REMARK 2. Let K be a finite extension of F and G be the group of all F -automorphism of K . Then the fixed field F_0 under G contains F , so $[K : F_0] \leq [K : F]$. Hence $o(G) \leq [K : F]$.

Definition 14.9. A finite extension K of a field F is said to be a *Galois Extension* of F if F is the fixed subfield of K under the group $G(K, F)$ of all F -automorphisms of K .

Corollary 14.10. Let $K = F(\alpha)$ be a simple finite separable extension of F . Then K is a splitting field of the minimal polynomial of α over F if and only if F is the fixed field under the group of all F -automorphisms of K .

Proof : Let $f(x)$ be the minimal polynomial of α over F and let $\deg f(x) = m$. Then $[K : F] = m$. Let $\alpha_1 = \alpha, \alpha_2, \alpha_3, \dots, \alpha_r$ be the distinct conjugates of α in K . Then $K = F(\alpha_i) \forall i = 1, 2, \dots, r$. For each i , there exists an F -automorphism σ_i of K such that $\sigma_i(\alpha_1) = \alpha_i$. Since α_1 generates K over F , each σ_i is uniquely determined. Further for any F -automorphism σ of K , as $\sigma(\alpha_1)$ is a conjugate of α_1 (Lemma 14.5), $\sigma(\alpha_1) = \alpha_i$ for some α_i . From this it follows that $\sigma = \sigma_i$. Hence the group $G(K, F)$ consists of $\sigma_1, \sigma_2, \dots, \sigma_r$. Let F_0 be the fixed field under $G(K, F)$. Then by the above theorem $[K : F_0] = o[G(K, F)] = r$. So $F = F_0$ if and only if $r = m$. Hence F is the fixed field under G if and only if $f(x)$ has all its m roots in K i.e., if and only if K is the splitting field of $f(x)$ over F .

EXAMPLE 4. Consider any prime integer p . Let $f(x) = x^{p-1} + x^{p-2} + \dots + x + 1$. Then

$$f(x+1) = x^{p-1} + {}^pC_1 x^{p-2} + {}^pC_2 x^{p-3} + \dots + {}^pC_{p-2} x + p.$$

Since p divides the coefficients of $x^{p-2}, x^{p-3}, \dots, x$ and constant term p ; p^2 does not divide the constant term and p does not divide the coefficient 1 of x^{p-1} , by Eisenstein's Criterion $f(x+1)$ is irreducible over $\mathbb{Q}$, hence $f(x)$ is irreducible.

$x^p - 1 = (x-1)(x^{p-1} + x^{p-2} + \dots + x + 1)$, so for any complex root α of $x^p - 1$ if $\alpha \neq 1$, then α is a root of $x^{p-1} + x^{p-2} + \dots + x + 1$.

Consider $\alpha_i = e^{\frac{2\pi i}{p}}, i = 1, 2, \dots, p-1$. None of α_i is equal to 1 and

all these α_i 's are distinct. Hence $\alpha_1, \alpha_2, \dots, \alpha_{p-1}$ are $p-1$ distinct roots of $f(x)$ over $\mathbf{Q}$. So $K = \mathbf{Q}(\alpha_1, \alpha_2, \dots, \alpha_{p-1})$ is the splitting field of the minimal polynomial $f(x)$ of α_1 over $\mathbf{Q}$. Hence the fixed field of the group G of all $\mathbf{Q}$ -automorphism of K is $\mathbf{Q}$. However $\mathbf{Q}$ is prime subfield, so every automorphism of K is a $\mathbf{Q}$ -automorphism. Hence $\mathbf{Q}$ is the fixed field under the group of all automorphisms of K . Thus K is a Galois extension of $\mathbf{Q}$.

EXAMPLE 5. Consider $K = \mathbf{Q}(\sqrt{-3}, \sqrt[3]{2})$. K is the splitting field of $x^3 - 2$ over $\mathbf{Q}$. Let G be the group of all automorphism of K . Indeed G is the group of all $\mathbf{Q}$ -automorphisms of K . Since $[K : \mathbf{Q}] = 6$ we have $o(G) \leq 6$ [by Remark 2 following theorem 14.8]. Suppose $x^3 - 2$ is not irreducible over $F = \mathbf{Q}(\sqrt{-3})$ then at least one root, say β , of $x^3 - 2$ is in $\mathbf{Q}(\sqrt{-3})$. So $\mathbf{Q}(\beta) \subseteq \mathbf{Q}(\sqrt{-3})$. Then $[\mathbf{Q}(\beta) : \mathbf{Q}] \mid [\mathbf{Q}(\sqrt{-3}) : \mathbf{Q}]$. However $[\mathbf{Q}(\beta) : \mathbf{Q}] = 3$ and $[\mathbf{Q}(\sqrt{-3}) : \mathbf{Q}] = 2$; we get 3 divides 2. That is not possible. Hence we get a contradiction.

Consequently $x^3 - 2$ is irreducible over F . Now $K = \mathbf{Q}(\sqrt{-3}, \sqrt[3]{2}) = F(\sqrt[3]{2})$ is a simple extension of F . As it contains all roots of $x^3 - 2$, K is the splitting field of the minimal polynomial $x^3 - 2$ of $\sqrt[3]{2}$ over F . Hence the group G_1 of all F -automorphisms of K is such that F is the fixed field under G_1 . So $3 = [K : F] = o(G_1)$. However G_1 is a subgroup of G . Consequently $3 \mid o(G)$. Again $x^2 + 3$ is irreducible over $F_1 = \mathbf{Q}(\sqrt[3]{2})$ and K is the splitting field of $x^2 + 3$ over F_1 with $[K : F_1] = 2$. So we also get $2 \mid o(G)$. Hence $6 \mid o(G)$. This shows $6 \leq o(G)$. Thus $o(G) = 6$. So it proves that $[K : \mathbf{Q}] = o(G)$. Consequently $\mathbf{Q}$ is the fixed field under G and K is a Galois extension of $\mathbf{Q}$.

EXAMPLE 5 (a). Consider the polynomial $x^4 - x^2 + 1$ over $\mathbf{Q}$. The splitting

field K of $x^4 - x^2 + 1$ over $\mathbf{Q}$ is given by $\mathbf{Q}\left(\frac{\pm\sqrt{3}\pm i}{2}\right) = \mathbf{Q}(\sqrt{3}, i)$.

Also $[K : \mathbf{Q}] = 4$. We determine the Galois group of K over $\mathbf{Q}$. Since $x^4 - x^2 + 1$ is irreducible over $\mathbf{Q}$ (Prove!), there exist $\sigma_1, \sigma_2, \sigma_3, \sigma_4 \in G(K, \mathbf{Q})$ such that $\sigma_1\left(\frac{\sqrt{3}+i}{2}\right) = \frac{\sqrt{3}+i}{2}$, $\sigma_2\left(\frac{\sqrt{3}+i}{2}\right) = \frac{\sqrt{3}-i}{2}$

$\sigma_3\left(\frac{\sqrt{3}+i}{2}\right) = \frac{-\sqrt{3}-i}{2}$, $\sigma_4\left(\frac{\sqrt{3}+i}{2}\right) = \frac{-\sqrt{3}+i}{2}$ Further for

any $\sigma \in G(K, \mathbf{Q})$, $\sigma\left(\frac{\sqrt{3}+i}{2}\right)$ is a root of $x^4 - x^2 + 1$. So $G(K, \mathbf{Q}) =$

$\{\sigma_1, \sigma_2, \sigma_3, \sigma_4\}$. Now $\sigma_1\left(\frac{\sqrt{3}+i}{2}\right) = \frac{\sqrt{3}+i}{2} \Rightarrow \sigma_1\left(\frac{\sqrt{3}-i}{2}\right) = \frac{2}{\sqrt{3}+i}$

$\sigma_2\left(\frac{\sqrt{3}+i}{2}\right) = \frac{\sqrt{3}-i}{2} \Rightarrow \sigma_2\left(\frac{\sqrt{3}-i}{2}\right) = \frac{2}{\sqrt{3}-i}$

$\sigma_3\left(\frac{\sqrt{3}+i}{2}\right) = \frac{-\sqrt{3}-i}{2} \Rightarrow \sigma_3\left(\frac{\sqrt{3}-i}{2}\right) = \frac{2}{-\sqrt{3}-i}$

$\sigma_4\left(\frac{\sqrt{3}+i}{2}\right) = \frac{-\sqrt{3}+i}{2} \Rightarrow \sigma_4\left(\frac{\sqrt{3}-i}{2}\right) = \frac{2}{-\sqrt{3}+i}$

$$= \left(\frac{\sqrt{3}+i}{2} \right)^{-1} = \frac{2}{\sqrt{3}+i} = \frac{\sqrt{3}-i}{2}, \quad \sigma_1 \left(\frac{-\sqrt{3}-i}{2} \right) = -\sigma_1 \left(\frac{\sqrt{3}+i}{2} \right) \\ = \frac{-\sqrt{3}-i}{2} \text{ and } \sigma_1 \left(\frac{-\sqrt{3}+i}{2} \right) = \frac{-\sqrt{3}+i}{2}.$$

This gives that $\sigma_1(\sqrt{3}) = \sigma_1 \left(\frac{\sqrt{3}+i}{2} + \frac{\sqrt{3}-i}{2} \right) = \frac{\sqrt{3}+i}{2} + \frac{\sqrt{3}-i}{2} = \sqrt{3}$, $\sigma_1(i) = \sigma_1 \left(\frac{\sqrt{3}+i}{2} - \frac{\sqrt{3}-i}{2} \right) = \frac{\sqrt{3}+i}{2} - \frac{\sqrt{3}-i}{2} = i$ and $\sigma_1(\sqrt{3}i) = \sigma_1(\sqrt{3})\sigma_1(i) = \sqrt{3}i$. But $\{1, \sqrt{3}, i, \sqrt{3}i\}$ is a basis of K over $\mathbb{Q}$. Hence $\sigma_1 = I$, the identity automorphism of σ . On similar lines it can be checked that $\sigma_2(\sqrt{3}) = \sqrt{3}$, $\sigma_2(i) = -i$, $\sigma_2(\sqrt{3}i) = -\sqrt{3}i$; $\sigma_3(\sqrt{3}) = -\sqrt{3}$, $\sigma_3(i) = -i$, $\sigma_3(\sqrt{3}i) = \sqrt{3}i$ and $\sigma_4(\sqrt{3}) = -\sqrt{3}$, $\sigma_4(i) = i$, $\sigma_4(\sqrt{3}i) = -\sqrt{3}i$. Now $G(K, \mathbb{Q})$ is a group of order 4 means that either it is a cyclic group of order 4 or it is isomorphic to Klein's 4-group.

Now $\sigma_2^2(\sqrt{3}) = \sigma_2(\sqrt{3}) = \sqrt{3}$, $\sigma_2^2(i) = \sigma_2(-i) = -\sigma_2(i) = i$ and $\sigma_2^2(\sqrt{3}i) = \sigma_2(-\sqrt{3}i) = -\sigma_2(\sqrt{3}i) = \sqrt{3}i \Rightarrow \sigma_2^2 = I$.

Again $\sigma_3^2(\sqrt{3}) = \sigma_3(-\sqrt{3}) = \sqrt{3}$, $\sigma_3^2(i) = \sigma_3(-i) = -\sigma_3(i) = i$, $\sigma_3^2(\sqrt{3}i) = \sigma_3(\sqrt{3}i) = \sqrt{3}i \Rightarrow \sigma_3^2 = I$. Similarly it can be verified that $\sigma_4^2 = I$. Hence $G(K, \mathbb{Q})$ has no element of order 4. This in turn implies that $G(K, \mathbb{Q})$ is isomorphic to Klein's 4-group. Thus the Galois group of K over $\mathbb{Q}$ is isomorphic to Klein's 4-group.

PROBLEMS

1. Verify all the postulates of a vector space for $M(S, K)$. If S is a finite set having n elements. Prove that $M(S, K)$ is of dimension n over K .

[Hint. For each $s \in S$. Define $f_s : S \rightarrow K$ by $f_s(s) = 1$ and $f_s(t) = 0 \forall t \neq s$. Show that these f_s 's constitute a basis of $M(S, K)$ over K .]

2. Let K be the splitting field of a polynomial $f(x)$ over a field F and let W be the set of all roots of $f(x)$ in K . For each $\tau \in G(K, F)$ let τ_1 be the restriction of τ to W . Show that τ_1 is a permutation of W . Further for any $\tau, \eta \in G(K, F)$, $\tau_1 = \eta_1$ if and only if $\tau = \eta$. Hence conclude that $G(K, F)$ is a group of permutations of W .

3. Let $K = F(a)$ be a simple algebraic extension of degree n of a field F of characteristic zero. Show that the number of conjugates of a in K over F divides $[K : F]$.

[Hint. If F_0 is the fixed field under $G(K, F)$, then $[K : F_0]$ is the number of conjugates of a in K over F .]

4. Let $K = F(a_1, a_2, \dots, a_n)$ be a finite algebraic extension of K . Show that each element of K is a linear combination over F of

monomials of the form $a_1 i_1 a_2 i_2 \dots a_n i_n$, $i_j \geq 0 \forall j=1, 2, \dots, n$. Hence deduce that any two F -automorphisms σ and η of K are equal if and only if they agree on the a_i 's i.e. $\sigma(a_i) = \eta(a_i) \forall i$.

5. Show that the identity automorphism is the only automorphism of a field having p elements, where p is a prime number.

6. Show that the Galois group of the splitting field K of $x^n - 1$ over $\mathbb{Q}$ is Abelian.

7. Show that in Example 5, the Galois group is isomorphic to S_3 .

8. Find the Galois group of the splitting field of each of the following polynomials over $\mathbb{Q}$.

(i) $x^3 - 2$

(ii) $x^4 - x^2 - 6$

(iii) $x^3 - 3$

(iv) $x^4 + 5x^2 + 6$

(v) $x^4 + 1$

(vi) $x^3 - x - 1$

(vii) $(x^2 - 3x + 1)^2(x^3 - 2)$

(viii) $x^4 + x^3 + x^2 + x + 1$

Ans. (i) Dihedral group of order 8 ((ii) Klein's 4-group, (iii) S_3 , (iv) Klein's 4 group, (v) Klein's 4-group, (vi) S_3 , (vii) $S_3 \times \mathbb{Z}_2$, (viii) $\mathbb{Z}_4$.

9. Find the Galois group of $x^6 - 2$ over $\mathbb{Q}(\sqrt{-3})$ [Ans. $\mathbb{Z}_6$]

10. Show that the Galois group of $(x^2 - 2)(x^3 - 3)$ over $\mathbb{Q}$ is isomorphic to $S_3 \times \mathbb{Z}_2$.

2. Normal Extensions and Fundamental Theorem of Galois Theory

Definition 14.11. An algebraic extension K of a field F is said to be a normal extension of F if each irreducible polynomial $f(x)$ over F , having a root in K , splits into linear factors over K .

In other words, we can say that an algebraic extension K of a field F is a normal extension of F if for each given $\beta \in K$ with $f(x)$ as its minimal polynomial of degree, say n , over F , the splitting field of $f(x)$ over F is contained in K .

Consider any field extension K of F of degree 2. Let $f(x)$ be any irreducible polynomial over F having a root, say α in K . Then $F(\alpha) \subseteq K$. So $[F(\alpha) : F] \leq 2$. However $[F(\alpha) : F] = \deg f(x)$. Hence $\deg f(x) = 1$ or $\deg f(x) = 2$. If $\deg f(x) = 1$, then $f(x)$ is of type $ax + b$, $a, b \in F$ and $a \neq 0$. Then the only root of $f(x)$ is $-b/a$, which is already in F and so F is the splitting field of $f(x)$ contained in K . If $\deg f(x) = 2$ then $f(x)$ is of the type $ax^2 + bx + c$; $a, b, c \in F$, $a \neq 0$. In that case if α is a root of $f(x)$ in K , we find $f(x) = a(x - \alpha)(x + \alpha + b/a)$

with $\alpha \in K$, $-(\alpha + b/a) \in K$. So $f(x)$ splits completely over K . This shows that K is a normal extension of F .

Now we give a theorem, which provides an excellent method of constructing large number of normal extensions.

Theorem 14.12. *Let K be a finite algebraic extension of a field F . Then K is a normal extension of F if and only if K is the splitting field over F of some non-zero polynomial over F .*

Proof : Let $K = F(a_1, a_2, \dots, a_n)$, a finite algebraic extension of F . Suppose K is a normal extension of F . For each a_i , let $f_i(x)$ be the minimal polynomial of a_i over F . Then $f_i(x)$ splits completely over K . Let $f(x) = f_1(x) \dots f_n(x)$. Now $f(x)$ also splits completely over K . So K contains a splitting field of $f(x)$. As $a_1, a_2, \dots, a_n$ are roots of $f(x)$ and K is generated by these roots of $f(x)$ over F , we conclude that K is also contained in the splitting field of $f(x)$ over F . So that K is equal to the splitting field of $f(x)$ over F .

Conversely let K be a splitting field of some polynomial $f(x) (\neq 0)$ over F . Let $a_1, a_2, \dots, a_n$ be the roots of $f(x)$ in K . Then $K = F(a_1, a_2, \dots, a_n)$. Let $p(x)$ be any irreducible polynomial over F having root β in K , $p(x)$ is also a polynomial over K but may not be irreducible. Let L be the splitting field of $p(x)$ over K . If we show that $L = K$, it will follow that $p(x)$ has all its roots in K i.e. $p(x)$ splits completely over K .

Suppose on the contrary that there exists a root β' of $p(x)$ in L such that $\beta' \notin K$. As β and β' are conjugate over F , there exists an F -isomorphism σ of $F(\beta)$ onto $F(\beta')$ such that $\sigma(\beta) = \beta'$. Now $F \subseteq F(\beta) \subseteq K$ gives K is also a splitting field of $f(x)$ over $F(\beta)$. Further $K(\beta') = F(a_1, a_2, \dots, a_n, \beta') = F(\beta') (a_1, a_2, \dots, a_n)$ gives $K(\beta')$ is a splitting field of $f(x)$ over $F(\beta')$. Hence there exists an isomorphism τ of K onto $K(\beta')$ such that $\sigma(x) = \tau(x) \forall x \in F(\beta)$, (Theorem 13.25) τ is also an F -isomorphism since σ is so. Let $f(x) = \alpha_0 + \alpha_1 x + \alpha_2 x^2 + \dots + \alpha_n x^n$, $\alpha_n \neq 0$ and $\alpha_i \in F$, $\forall i = 0, 1, 2, \dots, n$. Since $a_1, a_2, \dots, a_n$ are roots of $f(x)$ we get $f(x) = \alpha_n (x - a_1)(x - a_2) \dots (x - a_n)$. (1)

Then τ can be extended to τ' , the isomorphism of $K[x]$ onto $K(\beta')[x]$. Now $\tau' [f(x)] = \tau'(\alpha_0) + \tau'(\alpha_1)x' + \tau'(\alpha_2)x'^2 + \dots + \tau'(\alpha_n)x'^n = \alpha_0 + \alpha_1 x + \alpha_2 x^2 + \dots + \alpha_n x^n$ since $\tau'(x_i) = \tau(x_i) = \sigma(x_i) = \alpha_i$ for all i . This gives $\tau' [f(x)] = f(x)$.

Also $\tau' [f(x)] = \alpha_n [x - \tau(a_1)][x - \tau(a_2)] \dots [x - \tau(a_n)]$.
Hence $f(x) = \alpha_n [x - \tau(a_1)][x - \tau(a_2)] \dots [x - \tau(a_n)]$. (2)

From (1) and (2) we get that $\tau(a_1), \tau(a_2), \dots, \tau(a_n)$ are the same as $a_1, a_2, \dots, a_n$ arranged possibly in different order.

Thus $F(a_1, a_2, \dots, a_n) = F(\tau(a_1), \tau(a_2), \dots, \tau(a_n))$.

$$\begin{aligned}\text{However } K(\beta') &= \tau(K') = \tau[F(a_1, a_2, \dots, a_n)] \\ &= F[\tau(a_1), \tau(a_2), \dots, \tau(a_n)] \\ &= F(a_1, a_2, \dots, a_n) = K.\end{aligned}$$

Consequently $\beta' \in K$. This is a contradiction. So $p(x)$ splits completely over K . Hence K is normal extension of F . This proves the theorem. ■

Following are some corollaries of the above theorem.

Corollary 14.13. *Let K be a finite normal extension of a field F . If E is any subfield of K containing F then K is also a normal extension of E .*

Proof: Since K is a finite normal extension of F , there exists a polynomial $f(x)$ over F such that K is a splitting field of $f(x)$ over F . Then K is also a splitting field of $f(x)$ over E . Hence by the above theorem K is a normal extension of E . ■

Corollary 14.14. *Let K be a finite normal extension of a field F . If α_1, α_2 are any two elements of K conjugate over F , then there exists an F -automorphism σ of K such that $\sigma(\alpha_1) = \alpha_2$.*

Proof: Let K be the splitting field of a polynomial $f(x)$ over F . Now as α_1, α_2 are conjugates over F , there exists an F -isomorphism τ of $F(\alpha_1)$ onto $F(\alpha_2)$ such that $\tau(\alpha_1) = \alpha_2$. Now K is a splitting field of $f(x)$ over $F(\alpha_1)$ as well as $F(\alpha_2)$. So there exists an automorphism σ of K which extends τ i.e. $\sigma(x) = \tau(x) \forall x \in K$. Then $\sigma(\alpha_1) = \tau(\alpha_1) = \alpha_2$ and for any $\alpha \in F$, $\sigma(\alpha) = \tau(\alpha) = \alpha$. Hence σ is an F -automorphism of K such that $\sigma(\alpha_1) = \alpha_2$. ■

REMARK. Consider three fields $F \subseteq K \subseteq L$, such that K is a finite normal extension of F and L is finite normal extension of K . Can we say that L is normal extension of F ? The answer in general is no as can be shown by the following example.

EXAMPLE 6. Let $F = \mathbb{Q}$, $K = \mathbb{Q}(\sqrt{2})$, $L = \mathbb{Q}(\sqrt[4]{2})$. Since $[K : F] = 2$, K is a normal extension of F . $x^2 - \sqrt{2}$ is a polynomial over K such that $x^2 - \sqrt{2} = (x - \sqrt[4]{2})(x + \sqrt[4]{2})$ over L , a product of two linear factors and its roots generate L over K . So L is the splitting field of $x^2 - \sqrt{2}$ over K . Consequently L is normal extension of K . However $x^4 - 2$ is an irreducible polynomial over $\mathbb{Q}$, such that its one root $\sqrt[4]{2} \in L$, but another root $i\sqrt[4]{2}$ does not belong to L . Hence L is not a normal extension of $\mathbb{Q}$ i.e. F .

Much more can be said about normal extensions, we avoid to go into these aspects. Now let us recall that any algebraic extension of a field of characteristic zero, is a separable extension. However there exists fields of positive characteristic (in particular finite fields) whose all algebraic extensions are separable. Now we are going to

confine ourselves to finite normal extension of fields of characteristic zero. However it must be emphasized the majority of these results can be proved for finite normal and separable extensions. Following theorem is an apparent generalization of Corollary 14.10.

Theorem 14.15. Let K be a finite extension of a field F of characteristic zero. Then K is a normal extension of F if and only if the fixed field under the Galois group $G(K, F)$ is F itself. In case K is a normal extension of F , $[K : F] = o[G(K, F)]$.

Proof: Now $K = F(\alpha)$ for some $\alpha \in K$ (Theorem 13.50). Let K be a normal extension of F . So every irreducible polynomial over F having one root in K , splits into linear factors over K . Since α is algebraic over F we see that the splitting field K' of the minimal polynomial $f(x)$ of α over F , is contained in K . However $\alpha \in K'$ gives $F(\alpha) \subseteq K'$ i.e. $K \subseteq K'$. Hence $K = K'$. Thus F is the fixed field under $G(K, F)$, (Corollary 14.10).

Conversely let F be the fixed field under $G(K, F)$, then K is the splitting field of the minimal polynomial $f(x)$ of α over F . Hence K is normal extension of F (Theorem 14.12). This completes the theorem. Second part follows from Theorem 14.8. ■

Consider any finite normal extension K of a field F of characteristic zero. Let $G = G(K, F)$ and H , any subgroup of G . Let $K_H = \{x \in K \mid \sigma(x) = x \forall \sigma \in H\}$, the fixed field under H . For any subfield E of K containing F , $G(K, E) = \{\sigma \in G \mid \sigma(x) = x \forall x \in E\}$. $G(K, E)$ is the group of all E -automorphisms of K and is a subgroup of $G(K, F)$. Now we are in a position to prove the Fundamental Theorem of Galois theory (for the case of fields of characteristic zero only), giving an excellent duality between the family of all subfields of K containing F and the family of all subgroups of $G(K, F)$.

Theorem 14.16. (Fundamental Theorem of Galois Theory). Let K be a finite normal extension of a field F of characteristic zero and let $G(K, F)$ be the Galois group of K over F . Then the correspondence $E \leftrightarrow G(K, E)$ where E is a subfield of K containing F is 1-1 between the family of the subfields of K containing F and the family of all subgroups of $G(K, F)$, satisfying the following conditions.

- Given any subfield E of K containing F and subgroup H of $G(K, F)$
- (i) $E = K_{G(K, E)}$
 - (ii) $H = G(K, K_H)$
 - (iii) $[K : E] = o[G(K, E)]$ and $[E : F] = \text{index of } G(K, E) \text{ in } G(K, F)$
 - (iv) E is a normal extension of F if and only if $G(K, E)$ is a normal subgroup of $G(K, F)$.
 - (v) When E is a normal extension of F , then $G(E, F)$ is isomorphic to $G(K, F)/G(K, E)$.

Proof: Since K is a finite normal extension of F and $F \subseteq E \subseteq K$, we get that K is a finite normal extension of E . So E is same as the fixed field $K_{G(K, E)}$ (Theorem 14.15). Thus (i) follows.

By definition $K_H = \{x \in K \mid \sigma(x) = x \ \forall \sigma \in H\}$, each $\sigma \in H$ is a K_H -automorphism of K . So that $H \subseteq G(K, K_H)$. However $\sigma(H) = [K : K_H]$ (Theorem 14.8). At the same time as K is a normal extension of K_H , Theorem 14.15 gives that K_H is the fixed field under $G(K, K_H)$. So $[K : K_H] = o[G(K, K_H)]$. Thus $\sigma(H) = o[G(K, K_H)]$ and consequently $H = G(K, K_H)$. This proves (ii).

Now as K is a normal extension of F ; by Theorem 14.8, $[K : E] = o[G(K, E)]$. Thus $o[G(K, F)] = [K : F] = [K : E][E : F] = o[G(K, E)][E : F]$ gives $[E : F] = \frac{o[G(K, F)]}{o[G(K, E)]}$. This proves (iii).

Let E be a normal extension of F . Consider any $a \in E$, then the splitting field of the minimal polynomial of a over F is contained in E . That gives every conjugate of a over F in K is again in E . Since for any $\sigma \in G(K, F)$, $\sigma(a)$ is a conjugate of a , we have $\sigma(a) \in E$. Thus for any $\eta \in G(K, E)$, $\eta[\sigma(a)] = \sigma(a)$ and hence $(\sigma^{-1}\eta\sigma)(a) = a$. This proves $\sigma^{-1}\eta\sigma \in G(K, E)$ for every $\eta \in G(K, E)$ and $\sigma \in G(K, F)$. Consequently $G(K, E)$ is a normal subgroup of $G(K, F)$. Consider $a \in E$. As K is a normal extension of F , K contains a splitting field say L of the minimal polynomial $p(x)$ of a over F . Consider any root b of $p(x)$ in L . Then b is a conjugate of a over F . So there exists an F -automorphism σ of K such that $\sigma(a) = b$. For any $\eta \in G(K, E)$, $\sigma^{-1}\eta\sigma \in G(K, E)$ so $(\sigma^{-1}\eta\sigma)(a) = a$. Thus $\eta[\sigma(a)] = \sigma(a) \ \forall \eta \in G(K, E)$. Hence the fixed field under $G(K, E)$. This gives that $b = \sigma(a) \in E$. Hence $L \subseteq E$. This proves that E is a normal extension of F . Hence (iv) is proved.

Let E be a normal extension of F . Now $E = F(a)$ for some $a \in E$. For any $\sigma \in G(K, F)$, let σ_E denote the restriction of σ to E . Since $\sigma(a) \in E$, we get $\sigma(E) \subseteq E$. As $[\sigma(E) : F] = [E : F]$, we get $\sigma_E \in G(E, F)$. Define a mapping $\lambda : G(K, F) \rightarrow G(E, F)$ by $\lambda(\sigma) = \sigma_E \ \forall \sigma \in G(K, F)$. Clearly for any $\sigma, \eta \in G(K, F)$, $(\sigma\eta)_E = \sigma_E \eta_E$. Hence λ is a group homomorphism. Consider any $\gamma \in G(E, F)$. Now $\gamma(a)$ is a conjugate of a over F . Thus there exists an F -automorphism σ of K such that $\sigma(a) = \gamma(a)$. Further as σ and γ are both identity on F and E is generated by a over F we get $\sigma(x) = \gamma(x) \ \forall x \in F(a) = E$ i.e., $\gamma = \sigma_E = \lambda(\sigma)$. This proves λ is onto mapping. Hence $G(E, F) \cong \frac{G(K, F)}{\text{Ker } \lambda}$. Now $\sigma \in \text{Ker } \lambda$ if and only if σ_E is identity on E i.e., if and only if $\sigma(x) = x \ \forall$

$x \in E$ i.e., if and only if $\sigma \in G(K, E)$. Hence $\text{Ker } \lambda = G(K, E)$ and we obtain

$$G(E, F) \cong \frac{G(K, F)}{G(K, E)}.$$

This proves (v). Hence the theorem is proved. ■

In the following remarks K and F satisfy the hypothesis of the above theorem.

REMARK 1. Above theorem gives that the number of subfields of K containing F is finite, in fact this number is same as the number of subgroups of $G(K, F)$.

REMARK 2. Since every subgroup of an Abelian group is normal, above theorem gives that if $G(K, F)$ is Abelian, then every subfield E of K containing F is a normal extension of F . To illustrate such a situation, let $K = F(a)$ be such that all the conjugates of a are powers

of a . For example let $F = \mathbb{Q}$ and $K = \mathbb{Q}(a)$, where $a = e^{\frac{2\pi i}{p}}$ (See Example 4). Then $a, a^2, \dots, a^{p-1}$ are the roots of the irreducible polynomial $x^{p-1} + x^{p-2} + \dots + x + 1$. Consider any $\sigma, \eta \in G(K, F)$. As $\sigma(a)$ and $\eta(a)$ are both conjugate of a , we have $\sigma(a) = a^i, \eta(a) = a^j$ for some i, j . Then $(\eta\sigma)(a) = \eta[\sigma(a)] = \eta(a^i) = [\eta(a)]^i = (a^j)^i = a^{ij}$. Similarly $(\sigma\eta)(a) = a^{ji}$. Consequently $\sigma\eta = \eta\sigma$. This proves that $G(K, F)$ is an Abelian group.

Now we give an example to illustrate the use of Fundamental Theorem of Galois Theory.

EXAMPLE 6 (a). Consider $f(x) = x^4 - 5x^2 + 6$.

Since $f(x) = (x^2 - 3)(x^2 - 2)$; $\pm\sqrt{3}, \pm\sqrt{2}$ are roots of $f(x)$. Splitting field K of $f(x)$ over $\mathbb{Q}$ is $K = \mathbb{Q}(\pm\sqrt{2}, \pm\sqrt{3}) = \mathbb{Q}(\sqrt{2}, \sqrt{3})$. Now $\mathbb{Q} \subset \mathbb{Q}(\sqrt{2}) \subset \mathbb{Q}(\sqrt{2}, \sqrt{3})$. To prove last containment it is sufficient to show that $\sqrt{3} \notin \mathbb{Q}(\sqrt{2})$. Suppose this were not true, then $\sqrt{3} = a + b\sqrt{2}$ for some $a, b \in \mathbb{Q}$. This would in turn imply $3 = (a^2 + 2b^2) + 2\sqrt{2}ab \Rightarrow 2\sqrt{2}ab = 3 - a^2 - 2b^2$. Now $a = 0 \Rightarrow \sqrt{3} = b\sqrt{2} \Rightarrow \sqrt{6} = 2b \Rightarrow \sqrt{6}$ is rational! If $b = 0$ then $\sqrt{3}$ would be rational which is again absurd. Thus the only choice left is $\sqrt{2} = \frac{3 - a^2 - 2b^2}{2ab}$,

a rational number! Hence our supposition is wrong and as a consequence $\mathbb{Q}(\sqrt{2}, \sqrt{3}) \neq \mathbb{Q}(\sqrt{2})$. Similarly it can be shown that $\mathbb{Q} \subset \mathbb{Q}(\sqrt{3}) \subset \mathbb{Q}(\sqrt{2}, \sqrt{3})$.

Thus we see that $[K : \mathbb{Q}] = 4$. Then the Galois group $G(K, \mathbb{Q})$ being of order 4 is either cyclic or Klein's 4-group. Now K has at least two subfields $\mathbb{Q}(\sqrt{2})$ and $\mathbb{Q}(\sqrt{3})$ lying properly between $\mathbb{Q}$ and K , by Fundamental Theorem of Galois Theory $G(K, \mathbb{Q})$ has at least proper two subgroups lying between $\{I\}$ and itself. Since $[K : \mathbb{Q}(\sqrt{2})] = 2$ and $[K : \mathbb{Q}(\sqrt{3})] = 2$, $G(K, \mathbb{Q})$ has at least two subgroups of

order 2. Consequently $G(K, \mathbb{Q})$ cannot be cyclic, in other words $G(K, \mathbb{Q})$ is isomorphic to Klein's 4-group. As x^2-2 is irreducible over $\mathbb{Q}(\sqrt{3})$ so there exist automorphisms σ_1, σ_2 of $G[K, \mathbb{Q}(\sqrt{3})]$ such that $\sigma_1(\sqrt{2})=\sqrt{2}, \sigma_2(\sqrt{2})=-\sqrt{2}$. Similarly there exist automorphisms σ_3, σ_4 of $G[K, \mathbb{Q}(\sqrt{2})]$ such that $\sigma_3(\sqrt{3})=\sqrt{3}$ and $\sigma_4(\sqrt{3})=-\sqrt{3}$. It is easy to verify that there are four automorphisms in $G(K, \mathbb{Q})$ namely $\eta_1, \eta_2, \eta_3, \eta_4$ given by $\eta_1(\sqrt{2})=\sqrt{2}, \eta_1(\sqrt{3})=\sqrt{3}; \eta_2(\sqrt{2})=\sqrt{2}, \eta_2(\sqrt{3})=-\sqrt{3}; \eta_3(\sqrt{2})=-\sqrt{2}, \eta_3(\sqrt{3})=\sqrt{3}$ and $\eta_4(\sqrt{2})=-\sqrt{2}, \eta_4(\sqrt{3})=-\sqrt{3}$.

Using basis $\{1, \sqrt{2}, \sqrt{3}, \sqrt{2}\sqrt{3}\}$ of K over $\mathbb{Q}$, one can easily prove that $\eta_1=I, \eta_2^2=\eta_1^2=\eta_3^2=I$. Thus we have three subgroups $H_1=\{I, \eta_2\}; H_2=\{I, \eta_3\}; H_3=\{I, \eta_4\}$ of $G(K, \mathbb{Q})$. Corresponding fixed fields are determined as under.

Since $\eta_1(\sqrt{2})=\sqrt{2}$, every element of $\mathbb{Q}(\sqrt{2})$ is left fixed by η_2 . So $\mathbb{Q}(\sqrt{2}) \subseteq K_{H_1}$. Let $a+b\sqrt{2}+c\sqrt{3}+d\sqrt{2}\sqrt{3} \in K_{H_1}$ then $\eta_2(a+b\sqrt{2}+c\sqrt{3}+d\sqrt{2}\sqrt{3})=a+b\sqrt{2}+c\sqrt{3}+d\sqrt{2}\sqrt{3}$
 $\Rightarrow a+b\sqrt{2}-c\sqrt{3}-d\sqrt{2}\sqrt{3}=a+b\sqrt{2}+c\sqrt{3}+d\sqrt{2}\sqrt{3}$
 $\Rightarrow c\sqrt{3}+d\sqrt{2}\sqrt{3}=0 \Rightarrow c=d=0 \Rightarrow a+b\sqrt{2} \in \mathbb{Q}(\sqrt{2})$.

Hence $K_{H_1} \subseteq \mathbb{Q}(\sqrt{2})$. In other words $K_{H_1}=\mathbb{Q}(\sqrt{2})$. Similarly it can be checked that $K_{H_2}=\mathbb{Q}(\sqrt{3})$ and $K_{H_3}=\mathbb{Q}(\sqrt{2}\sqrt{3})$. Further each of H_1, H_2, H_3 is a normal subgroup of $G(K, \mathbb{Q})$. Also $[\mathbb{Q}(\sqrt{2}) : \mathbb{Q}]=2, [\mathbb{Q}(\sqrt{3}) : \mathbb{Q}]=2$ and $[\mathbb{Q}(\sqrt{2}\sqrt{3}) : \mathbb{Q}]=2 \Rightarrow \mathbb{Q}(\sqrt{2}), \mathbb{Q}(\sqrt{3}), \mathbb{Q}(\sqrt{2}\sqrt{3})$ are normal extension of $\mathbb{Q}$. Note that $H_1=G[K, \mathbb{Q}(\sqrt{2})], H_2=G[K, \mathbb{Q}(\sqrt{3})]$ and $H_3=G[K, \mathbb{Q}(\sqrt{2}\sqrt{3})]$.

Definition 14.16 (a). Let n be a positive integer. A complex number ω is called a *primitive n th root of unity* if $\omega^n=1$ but for no positive integer $m < n, \omega^m=1$.

REMARK. We know that all n th roots of unity form a cyclic group, say G , of order n under multiplication of complex numbers. Thus each generator of G is a primitive n th root of unity and vice versa.

EXAMPLE 6 (b). $\frac{-1+i\sqrt{3}}{2}$ is a primitive cube root of unity

and $\cos \frac{6\pi}{5} + i \sin \frac{6\pi}{5}$ is a primitive 5th root of unity. To check

these assertions, note that $\frac{-1+i\sqrt{3}}{2} = e^{\frac{2\pi i}{3}}$ and $\left(e^{\frac{2\pi i}{3}}\right)^3 = e^{2\pi i} = 1$ but $\left(e^{\frac{2\pi i}{3}}\right)^1 = \frac{-1+i\sqrt{3}}{2} \neq 1$ and $\left(e^{\frac{2\pi i}{3}}\right)^2 = e^{\frac{4\pi i}{3}} = \frac{-1-i\sqrt{3}}{2}$

$\neq 1$. Again $\left(e^{\frac{6\pi i}{5}}\right)^5 = 1$, but for no integer $0 < m < 5$, $\left(e^{\frac{6\pi i}{5}}\right)^m = 1$.

Definition 14.16 (b). Let S be the set of all primitive n th roots of unity, then $\phi_n(x) = \prod_{\xi \in S} (x - \xi)$ is called *cyclotomic polynomial* of degree n .

Firstly note that $e^{\frac{2\pi i}{n}}$ is a primitive n th root of unity and so, because of our earlier remark all the roots of unity are given by $1, \xi, \xi^2, \dots, \xi^{n-1}$ where $\xi = e^{\frac{2\pi i}{n}}$. So the splitting field of $x^n - 1$ over $\mathbf{Q}$ is $\mathbf{Q}(1, \xi, \xi^2, \dots, \xi^{n-1}) = \mathbf{Q}(\xi)$. By definition ξ satisfies $\phi_n(x)$. Once we prove that $\phi_n(x) \in \mathbf{Q}[x]$ and is irreducible over $\mathbf{Q}$, we shall be able to prove that $[\mathbf{Q}(\xi) : \mathbf{Q}] = \phi(n)$ (where $\phi(n)$ is Euler's totient function), since number of primitive n th roots of unity is same as the number of generators of a cyclic group of order n . Our next endeavour will be to prove all these assertions. Before proving next lemma, we note that $\phi_n(x)$ is a monic polynomial in $\mathbf{Q}(\xi)[x]$; ξ being a primitive n th root of unity.

Lemma 14.16 (c). $x^n - 1 = \prod_{\substack{d | n \\ 1 \leq d \leq n}} \phi_d(x)$.

Proof: As all n th roots of unity form a cyclic group G of order n , the order of any n th root of unity is a factor of n . Now $x^n - 1 = (x - \alpha_1)(x - \alpha_2) \dots (x - \alpha_n)$ where each α_i is an n th root of unity. We can put together those α_i 's whose orders are same. Since G is cyclic, there are α_i 's for each factor d of n . Suppose P is the product of all factors $(x - \alpha_i)$ of $x^n - 1$, such that $o(\alpha_i) = d$ and d is a factor of n . We claim that P is nothing but $\phi_d(x)$. Let α be a d th root of unity, α is also an n th root of unity, since $d | n \Rightarrow n = kd$ for some $k \in \mathbf{N} \Rightarrow \alpha^n = \alpha^{kd} = (\alpha^d)^k = 1$. In particular primitive d th roots of unity are n th roots of unity, so each one of the primitive d th root of unity must give rise to a factor in P . Hence $P = \phi_d(x)$. As a consequence

$$x^n - 1 = \prod_{\substack{d | n \\ 1 \leq d \leq n}} \phi_d(x). \blacksquare$$

This lemma gives a convenient method of determining explicitly $\phi_n(x)$ for small values of n . Thus $\phi_1(x) = x - 1$. Now $x^2 - 1 = \phi_1(x) \phi_2(x) \Rightarrow \phi_2(x) = \frac{x^2 - 1}{x - 1} = x + 1$. Again $x^3 - 1 = \phi_1(x) \phi_3(x) \Rightarrow \phi_3(x) = \frac{x^3 - 1}{x - 1} = x^2 + x + 1$. Finally $x^4 - 1 = \phi_1(x) \phi_2(x) \phi_4(x) \Rightarrow \phi_4(x) = \frac{x^4 - 1}{(x - 1)(x + 1)} = x^2 + 1$.

Theorem 14.16 (d). $\phi_n(x) \in \mathbb{Z}[x]$.

Proof: We apply induction on n . For $n=1$, $\phi_1(x) = x-1 \in \mathbb{Z}[x]$. Let now $n > 1$ and suppose $\phi_m[x] \in \mathbb{Z}[x]$ for all $1 \leq m < n$. By Lemma 14.16 (c), $x^n - 1 = \prod_{\substack{d|n \\ 1 \leq d < n}} \phi_d(x) = q(x) \phi_n(x)$ where $q(x) = \prod_{\substack{d|n \\ 1 \leq d < n}} \phi_d(x)$.

By induction hypothesis $\phi_d(x) \in \mathbb{Z}[x]$ for all $d < n$. Thus $q(x) \in \mathbb{Z}[x]$. Since $q(x)$ is monic polynomial in $\mathbb{Z}[x]$, by division algorithm in $\mathbb{Z}[x]$ we get $x^n - 1 = q(x)r(x) + s(x)$ where $r(x), s(x) \in \mathbb{Z}[x]$ and either $s(x) = 0$ or $\deg s(x) < \deg q(x)$. As $\mathbb{Z} \subseteq \mathbb{Q}(\xi)$, ξ is a primitive n th root of unity, we get that $x^n - 1 = q(x)r(x) + s(x)$ where $r(x), s(x) \in \mathbb{Q}(\xi)[x]$ with either $s(x) = 0$ or $\deg s(x) < \deg q(x)$. But in $\mathbb{Q}(\xi)[x]$; $x^n - 1 = q(x)\phi_n(x)$. Uniqueness of quotient and remainder implies that $s(x) = 0$ and $\phi_n(x) = r(x)$. Hence $\phi_n(x) \in \mathbb{Z}[x]$. ■

Corollary 14.16 (e). $\phi_n(x) \in \mathbb{Q}[x]$. ■

Finally we prove that $\phi_n(x)$ is irreducible over $\mathbb{Q}$.

Theorem 14.16 (f). $\phi_n(x)$ is irreducible over $\mathbb{Q}$.

Proof: In the light of Lemma 10.41 it is sufficient to prove that $\phi_n(x)$ is irreducible over $\mathbb{Z}$. Let $h(x)$ be an irreducible polynomial $\in \mathbb{Z}[x]$ such that $\phi_n(x) = h(x)g(x)$ for some $g(x) \in \mathbb{Z}[x]$. Since $\phi_n(x)$ is monic, both $h(x)$ and $g(x)$ can be chosen to be monic. Now any root of $h(x)$ is a root of $\phi_n(x)$, i.e., a primitive n th root of unity. Let ξ be a primitive n th root of unity so that $h(\xi) = 0$. Let p be a prime number such that $p < n$ and $p \nmid n$. Firstly we claim that ξ^p is a root of $h(x)$. Suppose that ξ^p is not a root of $h(x)$. Now $(p, n) = 1 \Rightarrow \xi^p$ is a primitive n th root of unity $\Rightarrow \phi_n(\xi^p) = 0 \Rightarrow h(\xi^p)g(\xi^p) = 0 \Rightarrow g(\xi^p) = 0$ as $h(\xi^p) \neq 0$. Hence ξ^p is a root of $g(x)$; in other words ξ is a root of $g(x^p)$. But $h(x)$ is an irreducible monic polynomial satisfied by ξ , so it is the minimal polynomial of ξ over $\mathbb{Q}$. Hence $h(x) \mid g(x^p)$ in $\mathbb{Q}[x]$. This in turn implies that $g(x^p) = h(x)l(x)$. Since $g(x^p), h(x) \in \mathbb{Z}[x]$, $l(x) \in \mathbb{Z}[x]$, (use same arguments as are given in Theorem 14.16 (d)). Thus $x^n - 1 = q(x)\phi_n(x)$ where $q(x) = \prod_{\substack{d|n \\ 1 \leq d < n}} \phi_d(x) \Rightarrow x^n - 1 = q(x)h(x)g(x)$.

Now define $\theta : \mathbb{Z} \rightarrow \mathbb{Z}/\langle p \rangle$ by $\theta(n) = n + \langle p \rangle = \bar{n}$. θ is an onto homomorphism. θ induces an onto homomorphism $\phi : \mathbb{Z}[x] \rightarrow \frac{\mathbb{Z}}{\langle p \rangle}[x]$

given by $\phi(\sum_{i=1}^t \alpha_i x^i) = \sum_{i=1}^t \bar{\alpha}_i x^i$. Let us for convenience sake denote

$\phi(f(x))$ by $\bar{f}(x)$ for any $f(x) \in \mathbb{Z}[x]$ (1)

$$\phi(x^n - 1) = \phi[q(x)h(x)g(x)] \Rightarrow x^n - \bar{1} = \bar{q}(x)\bar{h}(x)\bar{g}(x)$$

Let $g(x) = \beta_0 + \beta_1 x + \dots + \beta_r x^r$.
By Fermat's Theorem (Problem 30, § 2, Chapter 2) for $a \in \mathbb{Z}$,

$\bar{a}^p = \bar{a}$. This implies that $\bar{g}(x^p) = (\bar{\beta}_0 + \bar{\beta}_1 x^p + \dots + \bar{\beta}_r x^{rp}) = \bar{\beta}_0^p + (\bar{\beta}_1^p x^p + \dots + \bar{\beta}_r^p x^{rp}) = (\bar{\beta}_0 + \bar{\beta}_1 x + \dots + \bar{\beta}_r x^r)^p = [\bar{g}(x)]^p$. Now $g(x^p) = h(x)l(x) \Rightarrow g(x^p) = \bar{h}(x)\bar{l}(x) \Rightarrow [\bar{g}(x)]^p = \bar{h}(x)\bar{l}(x)$. If $\bar{l}(x)$ is irreducible factor of $\bar{h}(x)$ in $\frac{\mathbf{Z}}{\langle p \rangle}[x]$ then $\bar{l}(x) \mid [\bar{g}(x)]^p$ in $\frac{\mathbf{Z}}{\langle p \rangle}[x] \Rightarrow \bar{l}(x) \mid \bar{g}(x)$ in $\frac{\mathbf{Z}}{\langle p \rangle}[x]$. Using (1) we see that $[\bar{l}(x)]^p \mid x^n - \bar{1}$ in $\frac{\mathbf{Z}}{\langle p \rangle}[x]$. Consequently $x^n - \bar{1}$ is not separable over $\mathbf{Z}/\langle p \rangle$. But $f(x) = x^n - \bar{1} \Rightarrow [f(x)]' = nx^{n-1} \neq 0$ as $p \nmid n$. This is a contradiction to Lemma 13.41. Hence our claim that ξ^p is a root of $h(x)$ is established. Let ζ be any primitive n th root of unity and since ξ is also a primitive n th root of unity we get that $\zeta = \xi^k$ for some $k \in \mathbf{N}$. Let $k = p_1 p_2 \dots p_s$ where p_i 's ($1 \leq i \leq s$) are prime numbers (not necessarily distinct) such that $p_i < n \forall i$. None of p_i 's is a factor of n . Otherwise $(k, n) \neq 1$, i.e., $\zeta = \xi^k$ can't be a primitive n th root of unity! By successive application of what we have proved ξ^{p_1} ; $(\xi^{p_1})^{p_2}$; $(\xi^{p_1 p_2})^{p_3}$; $\dots$; $(\xi^{p_1 p_2 \dots p_{s-1}})^{p_s} = \zeta$ are roots of $h(x)$. Thus every root of $\phi_n(x)$ is a root of $h(x)$. Consequently $\phi_n(x) = h(x)$. Hence $\phi_n(x)$ is irreducible over $\mathbf{Z}$. ■

Using Corollary 14.16 (e) and Theorem 14.16 (f) we get the following corollary.

Corollary 14.16 (g). The degree of splitting field of $x^n - 1$ over $\mathbf{Q}$ is $\phi(n)$ ■

PROBLEMS

1. Let F be a Galois field having q elements and K be a field extension of F of degree n . Prove that K is a normal extension of F .

[Hint. K is a splitting field over F of the polynomial $x^{q^n} - x$].

2. Let K be any algebraic field extension of F . If $\{F_i\}_{i \in I}$ be any non-void family of subfields containing F , such that each F_i is a normal extension of F . Show that the intersection $\bigcap_{i \in I} F_i$ is also a normal extension of F .

3. Let $F \subseteq K \subseteq L$ be three fields such that F is of characteristic zero, K is a finite normal extension of F and L is a finite normal extension of K . Prove that following:

If σ is any F -automorphism of K then there exists an F -automorphism η of L such that σ is restriction of η to K .

[Hint. Let $K = F(a)$, $L = K(b)$. For any $\sigma \in G(K, F)$, $\sigma(a)$ is a conjugate of a over F with $F(a) = F(\sigma(a))$. Further L is the splitting field of the minimal polynomial of b over K as well as $F(\sigma(a))$.]

4. Let K be a finite extension of a field F . Prove the following:

(i) There exists a minimal normal extension of F containing K .

[Hint. Let $K = F(a_1, a_2, \dots, a_n)$. For each i let $f_i(x)$ be the minimal polynomial of a_i over F and let $f(x)$ be the product of these $f_i(x)$'s. Then the splitting field L of $f(x)$ over K is the desired field].

(ii) K is a normal extension of F if and only if it satisfies the following condition: For every extension K' of K , any F -isomorphism of K in K' is essentially an F -automorphism of K itself.

5. Give example of a normal extension of $\mathbb{Q}$ of degree 3 and an example of an algebraic extension of $\mathbb{Q}$ of degree 3 which is not a normal extension.

[Hint. Consider the splitting field of the polynomial $x^3 - 3x + 1$ over $\mathbb{Q}$].

6. Let $f(x)$ be any non-zero polynomial over a field F and let K be the splitting field of $f(x)$ over F . Then the group $G(K, F)$ is called the *Galois group* of $f(x)$. Prove that the Galois group G of the polynomial $p_n(x) = x^n - 1$ over the rational numbers is cyclic.

[Hint. Notice that $K = \mathbb{Q}(\alpha)$ where $\alpha = e^{\frac{2\pi i}{n}}$ is the splitting field of $x^n - 1$ over $\mathbb{Q}$ and every conjugate of α is a power of α .]

7. Let K be a finite normal extension of a field F of characteristic zero with $[K : F] = n$. Let $\sigma_1, \sigma_2, \dots, \sigma_n$ be n F -automorphisms of K . For each $x \in K$, define $N_{K/F}(x)$, $T_{K/F}(x)$ (called the *norm* and *trace* of x over F respectively) as follows.

$$N_{K/F}(x) = \prod_{i=1}^n \sigma_i(x)$$

$$T_{K/F}(x) = \sum_{i=1}^n \sigma_i(x),$$

and

Prove the following:

(i) For any F -automorphism σ of K
 $\sigma[N_{K/F}(x)] = N_{K/F}(x)$, $\sigma[T_{K/F}(x)] = T_{K/F}(x)$

Hence or otherwise deduce that $N_{K/F}(x)$ and $T_{K/F}(x)$ are both in F .
 (ii) Consider $K = \mathbb{Q}(\sqrt{-3}, \sqrt[3]{2})$. Compute $N_{K/\mathbb{Q}}(\sqrt[3]{2})$ and $T_{K/\mathbb{Q}}(\sqrt[3]{2})$.

(iii) If E is any subfield of K containing F such that E is a normal extension of F , then for any $x \in K$.

$$N_{K/F}(x) = N_{E/F}[N_{K/E}(x)]$$

$$T_{K/F}(x) = T_{E/F}[T_{K/E}(x)].$$

and

8. Let K be a finite extension of a field F . Prove that K is a

separable normal extension of F if and only if F is the fixed field under the group $G(K, F)$.

9. Using problem 8, prove the Fundamental Theorem of Galois Theory under the assumption that K is a separable normal extension of F of finite degree.

10. Let K be a normal extension of a field F of characteristic zero such that $[K : F] = 4$. Prove that every subfield of K containing F is a normal extension of F .

[Hint. Any group of order 4 is Abelian].

3. Radical Extensions and Solvability by Radicals

In this and next sections we are going to discuss some simple but fundamental applications of field theory developed so far, to the Theory of Equations and Geometry. Before we actually take up the concept of solvability of a polynomial by radicals, we prove some more results on general theory of fields.

Lemma 14.17. Any finite subgroup of the multiplicative group of a field is cyclic.

Proof: Let F be a field and S be any finite subgroup of the multiplicative group F^* of F . Since S is finite we can find an element $x \in S$ having largest order (with respect to multiplication) among the elements of S . Then for every $y \in S$, $o(y) \mid o(x)$ (Lemma 4.25). Thus if $o(x) = n$ then $y^n = e$ for all $y \in S$ where e is the unity of F . Consequently every element of S is a root of the polynomial $ex^n - e$. Now $e, x, x^2, \dots, x^{n-1}$ are n distinct elements of S . Since a polynomial of degree n over a field cannot have more than n roots, we get $e, x, x^2, \dots, x^{n-1}$ are the only elements of S . Thus $S = \langle x \rangle$, a cyclic group. ■ [c.f. Theorem 13.56].

In particular as F^* is its own subgroup, we see that if F^* is finite then it must be cyclic. F^* is finite whenever F is a Galois Field. So we get the following.

Corollary 14.18. The multiplicative group of any Galois field is cyclic. ■

Definition 14.19. For a positive integer n , an element ζ of a field F is called an n th root of unity if $\zeta^n = e$ where e is the unity of F .

REMARK. Thus each n th root of unity is a root of the polynomial $f(x) = ex^n - e$. Now $f'(x) = nex^{n-1}$. Let the field F be such that the characteristic of F does not divide n . $f(x)$ is a polynomial over F . Since the characteristic of F does not divide n , $f'(x) = nex^{n-1} \neq 0$. Now only possible root of $f'(x)$ is $x = 0$ and 0 is not a root of $f(x)$. That gives $f(x)$ and $f'(x)$ can not have a common root in any field extension K of F . Consequently $f(x)$ has

no repeated root in any field extension of F (Lemma 13.41). So in particular if K is the splitting field of $f(x)$ over F then K contains n distinct roots of $f(x)$.

Theorem 14.20. Let F be a field of characteristic zero and K be the splitting field of the polynomial $ex^n - e$ over F , then

(i) the set of all roots of $f(x)$ is a cyclic group of order n with respect to multiplication.

(iii) the Galois group $G(K, F)$ is isomorphic to a subgroup of the residue class group M of integers relatively prime to n . Consequently $G(K, F)$ is a commutative group.

Proof: (i) Let S be the set of all roots of $ex^n - e$ in K . Then S has n members (see the remark above). Let ζ and $\lambda \in S$. Then $\zeta^n = e$, $\lambda^n = e$. This gives that $(\zeta\lambda^{-1})^n = \zeta^n(\lambda^n)^{-1} = ee^{-1} = e$, $\zeta\lambda^{-1}$ is also a root of $ex^n - e$. Thus $\zeta\eta^{-1} \in S$. Hence S is a finite subgroup of the multiplicative group of K . Consequently S is a cyclic group of order n (Lemma 14.17).

(ii) Let θ be a generator of the cyclic group S of all the n roots of $ex^n - e$. Then θ is of order n with respect to multiplication. Hence θ is a primitive n th root of unity. Now every other member of S is a power of θ . Consequently $K = F(\theta)$. Let σ be any F -automorphism of K . For any n th root ζ of unity $\zeta^n = e$ gives $e = \sigma(e) = \sigma(\zeta^n) = [\sigma(\zeta)]^n$. So $\sigma(\zeta) \in S$. Since σ is 1-1 we see that the restriction of σ to S is an automorphism of the group $S = \langle \theta \rangle$. Consequently $\sigma(\theta)$ also generates S . So if $\sigma(\theta) = \theta^t$ then t is an integer relatively prime to n . Let $\bar{t}$ be the residue class modulo n to which t belongs, then $\bar{t} \in M$. Define a mapping

$g : G(K, F) \rightarrow M$
by putting $g(\sigma) = \bar{t}$ whenever $\sigma(\theta) = \theta^t$. Firstly we claim that g is well-defined. For let also $\sigma(\theta) = \theta^u$, then $\theta^{t-u} = e$. Consequently $n \mid (t-u)$. So $\bar{t} = \bar{u}$. For any two $\sigma, \eta \in G(K, F)$; let $g(\sigma) = \bar{t}_1$, $g(\eta) = \bar{t}_2$. Then $\sigma(\theta) = \theta^{t_1}$, $\eta(\theta) = \theta^{t_2}$. Consequently $(\sigma\eta)(\theta) = \sigma[\eta(\theta)] = \sigma(\theta^{t_2}) = [\sigma(\theta)]^{t_2} = (\theta^{t_1})^{t_2} = \theta^{t_1 t_2}$. Thus $g(\sigma\eta) = \overline{t_1 t_2} = \bar{t}_1 \bar{t}_2 = g(\sigma)g(\eta)$, showing that g is a group homomorphism. To show that g is 1-1, let $\sigma \in \text{Ker } g = [\sigma(\theta)]^{t_2} = (\theta^{t_1})^{t_2} = \theta^{t_1 t_2}$. Thus $g(\sigma\eta) = \overline{t_1 t_2} = \bar{t}_1 \bar{t}_2 = g(\sigma)g(\eta)$, showing that g is a group homomorphism. To show that g is 1-1, let $\sigma \in \text{Ker } g$ then $g(\sigma) = \bar{1}$, so $\sigma(\theta) = \theta^1 = \theta$. As σ is identity on F and $K = F(\theta)$ we get σ is identity on K . Hence $\text{Ker } g = (1)$. This proves (ii). ■

REMARK 1. Note that in the proof of part (ii), we do not need impose any restriction on the characteristic of F .

REMARK 2. Let F be a field of characteristic zero and $\alpha (\neq 0) \in F$. Let K be a field extension of F containing the splitting field of the polynomial $ex^n - \alpha$ and also a primitive n th root of unity say θ . Then $e, \theta, \theta^2, \dots, \theta^{n-1}$ are the totality of n th roots of unity. Let a be any arbitrary root of the polynomial $ex^n - \alpha$ in K . Then for each θ^i , $i=0$,

$1, \dots, n-1$; $(\theta^i a)^n = \theta^{ni} a^n = \epsilon a = \alpha$ gives that each of $\theta^i a$ is a root of $ex^n - \alpha$. Thus the totality of n roots of $ex^n - \alpha$ is given by $a, \theta a, \theta^2 a, \dots, \theta^{n-1} a$. Let L be the splitting field of $ex^n - \alpha$ over F contained in K . Then $a, \theta a, \theta^2 a, \dots, \theta^{n-1} a \in L$. Hence $F(\theta, a) \subseteq L$. But $L = F(a, \theta a, \theta^2 a, \dots, \theta^{n-1} a) \subseteq F(\theta, a)$. Hence $L = F(\theta, a)$.

In particular if F contains a primitive n th root of unity then $\theta \in F$ and $L = F(a)$. So one can also say that if F contains a primitive n th root of unity then for any $\alpha (\neq 0) \in F$, the splitting field L of $ex^n - \alpha$ over F is $F(a)$ where a is any arbitrary root of $ex^n - \alpha$.

Lemma 14.21. *If a field F of characteristic zero contains all the n th roots of unity 1 (for some fixed positive integer n) then for any $\alpha (\neq 0) \in F$, the splitting field L of $x^n - \alpha$ is given by $L = F(a)$ where a is an arbitrary root of $x^n - \alpha$, further the Galois Group $G(L, F)$ is Abelian.*

Proof: Let ζ be a primitive n th root of unity in F and let q be any root of $x^n - \alpha$ in the splitting field L of $x^n - \alpha$ over F . As seen above (Remark 2) $L = F(a)$ and the roots of $x^n - \alpha$ in L are $a, \zeta a, \zeta^2 a, \dots, \zeta^{n-1} a$. Consider any two members σ and η of $G(L, F)$. Now $\sigma(a)$ and $\eta(a)$ are both roots of $x^n - \alpha$. Consequently $\sigma(a) = \zeta^i a$, $\eta(a) = \zeta^j a$ for some i and j . Then $(\sigma\eta)(a) = \sigma[\eta(a)] = \sigma(\zeta^j a) = \zeta^j \sigma(a)$ as σ is an F automorphism so $\sigma\eta(a) = \zeta^j \zeta^i a = \zeta^{i+j} a$. Similarly $\eta\sigma(a) = \zeta^{i+j} a$, so that $\sigma\eta(a) = \eta\sigma(a)$. Consequently $\sigma\eta = \eta\sigma$. Hence $G(L, F)$ is an Abelian group. ■

Definition 14.22. A field extension K of a field F is said to be a **radical extension** of F if there exists a sequence of fields

$F = F_0 \subseteq F_1 \subseteq F_2 \subseteq \dots \subseteq F_r = K$
such that for each $i = 1, 2, \dots, r$; $F_i = F_{i-1}(\omega_i)$ for some $\omega_i \in F_i$ with the property that $\omega_i^{r_i} \in F_{i-1}$ for some $r_i \geq 1$.

REMARK. Thus one notices that $K = F(\omega_1, \omega_2, \dots, \omega_r)$, $F_i = F(\omega_1, \omega_2, \dots, \omega_i)$ and $\omega_i^{r_i} \in F_{i-1}$. Now $\omega_i^{r_i} \in F_{i-1}$ gives that ω_i is a root of the polynomial $x^{r_i} - \omega_i^{r_i}$ over F_{i-1} . So that F_i is a simple algebraic extension of F_{i-1} and consequently $[F_i : F_{i-1}]$ is finite. Then $[K : F] = [F_r : F_{r-1}][F_{r-1} : F_{r-2}] \dots [F_1 : F]$ gives that K itself is a finite algebraic extension of F .

Now we give a precise definition of what we mean by saying that a polynomial is solvable by radicals.

Definition 14.23. A non constant polynomial $p(x)$ over a field F is said to be **solvable by radicals** if the splitting field K of $p(x)$ over F is contained in a radical extension of F .

EXAMPLE 7. Let $f(x) = x^2 + ax + b$ be a monic quadratic polynomial over $\mathbb{Q}$, the field of rational numbers. Its two roots are

$\alpha_1 = \frac{-a + \sqrt{a^2 - 4b}}{2}$, $\alpha_2 = \frac{-a - \sqrt{a^2 - 4b}}{2}$. If we take $L = \mathbb{Q}(\omega)$ where

$\omega = \sqrt{a^2 - 4b}$ then we see that $\omega \in \mathbf{Q}$. So that L is a radical extension of $\mathbf{Q}$. Further L itself is the splitting field of $f(x)$ over $\mathbf{Q}$. Hence $f(x)$ is solvable by radicals.

EXAMPLE 8. Consider the cubic $f(x) = x^3 + 3ax^2 + 3bx + c$ over $\mathbf{Q}$, the field of rational numbers. If we put $z = x + a$ then the above equation becomes an equation of the form $g(z) = z^3 + 3b_1z + c_1$, $b_1, c_1 \in \mathbf{Q}$. If we know the roots of $g(z)$ then the roots of $f(x)$ are obtained from those of $g(z)$ by subtracting a from each root of $g(z)$. Since $a \in \mathbf{Q}$, we see that the splitting fields of $f(x)$ and the corresponding $g(z)$ are same. Cardan's formula gives that the roots of $g(z)$ are $p + q, \omega p + \omega^2 q, \omega^2 p + \omega q$ where ω is imaginary cube root of unity,

$$p = \sqrt[3]{-\frac{c_1}{2} + \sqrt{b_1^3 + \frac{c_1^2}{4}}}$$

$$\text{and } q = \sqrt[3]{-\frac{c_1}{2} + \sqrt{b_1^3 + \frac{c_1^2}{4}}}$$

$$\text{Take } \alpha_1 = \sqrt[3]{b_1^3 + \frac{c_1^2}{4}}, \alpha_2 = \sqrt[3]{-\frac{c_1}{2} + \alpha_1}, \alpha_3 = \sqrt{-3}.$$

Now take $F_0 = \mathbf{Q}$, $F_1 = F_0(\alpha_1)$, $F_2 = F_1(\alpha_2)$, $F_3 = F_2(\alpha_3)$. We find that $F_0 < F_1 < F_2 < F_3$; $\alpha_1^2 \in F_0$, $\alpha_2^3 \in F_1$, $\alpha_3^2 \in F_2$. Further F_3 contains all the roots of $g(z)$. Hence $g(z)$ and consequently $f(x)$ is solvable by radicals.

REMARK. For any polynomial $f(x) = x^n + a_1x^{n-1} + a_2x^{n-2} + \dots + a_{n-1}x + a_n$ over a field F , we can say that $f(x)$ is solvable by radicals if we can obtain any root of $f(x)$ by applying finite number of operations of addition, multiplication, subtraction, division by non-zero elements and taking m th root (for any positive integer m) on the coefficients of $f(x)$.

Theorem 14.23. If a field E is a finite normal extension of F by radicals then there exists a finite normal extension K of F containing E such that K is an extension of F by radicals.

Proof: Since E is an extension of F by radicals, there exists a chain of fields $F = F_0 \subseteq F_1 \subseteq F_2 \subseteq F_3 \subseteq \dots \subseteq F_k = E$ such that for each $i = 1, 2, \dots, k$; $F_i = F_{i-1}(\omega_i)$ for some $\omega_i \in F_i$ with $\omega_i^{r_i} \in F_{i-1}$ for some integer $r_i \geq 1$. Now each of these ω_i 's is algebraic over F . Let $f_i(x)$ be the minimal polynomial of ω_i over F and let $f(x) = f_1(x)f_2(x)\dots f_k(x)$.

Let K be the splitting field of $f(x)$ over E . Then $E \subseteq K$ and as $\omega_1, \omega_2, \dots, \omega_k$ are also roots of $f(x)$ we get that K is the splitting field of $f(x)$ over F . Thus K is a normal extension of F of finite degree. This gives that K has only finite number $\sigma_0, \sigma_1, \dots, \sigma_m$ of F automorphisms

with $\sigma_0 = I$, the identity automorphism of K .

Define $b_{tk+i} = \sigma_t(\omega_i) \forall t=0, 1, 2, \dots, m$ and $i=1, 2, \dots, k$. Since $f_i(x)$ is the minimal polynomial of ω_i over F and σ_t is an F -automorphism of K , $\sigma_t(\omega_i)$ and ω_i are conjugates over F . So each of b_{tk+i} is a root of $f_i(x)$ and hence also of $f(x)$. Consider any root c of $f(x)$, then c is a root of some $f_i(x)$. As ω_i and c are roots of same polynomial $f_i(x)$ irreducible over F there exists an F -automorphism σ_t of K such that $\sigma_t(\omega_i) = c$. Thus, by definition $c = b_{tk+i}$. This, establishes that the only roots of $f(x)$ are $b_1, b_2, \dots, b_k, b_{k+1}, \dots, b_{m+k}$. This in turn implies that $K = F(b_1, b_2, \dots, b_{m+k})$. Since for any $s=1, 2, \dots, b_{tk+s} = \sigma_t(\omega_s)$, we get for each $l=1, 2, 3, \dots, k$;

$$\begin{aligned}\sigma_t F(\omega_1, \omega_2, \dots, \omega_l) &= \sigma_t(F_l) \\ &= F(\sigma_t(\omega_1), \sigma_t(\omega_2), \dots, \sigma_t(\omega_l)) \\ &= F(b_{tk+1}, b_{tk+2}, \dots, b_{tk+l}).\end{aligned}$$

As $\omega_l^{r_l} \in F_{l-1} = F(\omega_1, \omega_2, \dots, \omega_{l-1})$, we get $b_{tk+l}^{r_l} = [\sigma_t(\omega_l)]^{r_l}$
 $= \sigma_t(\omega_l^{r_l}) \in \sigma_t(F_{l-1}) = F(b_{tk+1}, \dots, b_{tk+l-1})$.

If we define $K_{tk+l} = F(b_1, b_2, \dots, b_{tk+l})$ for all $t=0, 1, 2, \dots, m$ and $l=1, 2, \dots, k$; we have $K_{tk+l} = K_{tk+l-1}(b_{tk+l})$, such that
 $K_1 \subseteq K_2 \subseteq K_3 \subseteq \dots \subseteq K_k \subseteq K_{k+1} \subseteq \dots \subseteq K_{2k} \subseteq \dots \subseteq K_{mk} \subseteq K_{m+k} = K$. (1)

Further $b_{tk+l}^{r_l} = [\sigma_t(\omega_l)]^{r_l}$

$$\begin{aligned}&= \sigma_t(\omega_l^{r_l}) \in \sigma_t(F_{l-1}) = F(b_{tk+1}, \dots, b_{tk+l-1}) \\ &\subseteq F(b_1, b_2, \dots, b_{tk+l-1}) = K_{tk+l-1}.\end{aligned}$$

From this result and sequence (1) of fields we get that K is a radical extension of F . Hence K is a normal extension of F containing E such that K is also a radical extension of F . ■

Theorem 14.24. Let E be a normal extension of a field F of degree n . Further let F have a primitive n th root of unity. Then the Galois group $G(E, F)$ is cyclic if and only if there exists $b \in F$ such that $x^n - b$ is irreducible over F and E is the splitting field of $x^n - b$ over F .

Proof: Let for some positive integer n and for some $b \in F$; $x^n - b$ be irreducible over F and let E be the splitting field of this polynomial over F . Then $E = F(c)$ (Lemma 14.21) where c is any root of $x^n - b$. Thus $[E : F] = n$ Hence $G(E, F)$ is of order n . Now let ξ be a primitive n th root of unity in F (Note that if F contains a primitive n th root of unity then it contains all primitive n th roots of unity!). Then $c, \xi c, \xi^2 c, \dots, \xi^{n-1} c$ are the n roots of $x^n - b$ (Remark before Lemma 14.21). Since $\xi, \xi c$ are conjugates over F there exists an F -isomorphism σ of $F(c)$ onto $F(\xi c)$ such that $\sigma(c) = \xi c$. However

$E = F(c) = F(\xi c)$.² So σ is an F -automorphism of E . Thus σ is in $G(E, F)$. Now $\sigma^k(c) = \sigma[\sigma(c)] = \sigma(\xi c) = \xi \sigma(c) = \xi^2 c$. In general $\sigma^k(c) = \xi^k c \forall k = 1, 2, \dots, n$. As $\xi^n = 1$ and $c, \xi c, \dots, \xi^{n-1}c$ are all distinct we get that $I, \sigma, \sigma^2, \dots, \sigma^{n-1}$ are all n distinct members of $G(E, F)$. Hence $G(E, F) = \langle \sigma \rangle$, a cyclic group order n .

Conversely let $G(E, F)$ be a cyclic group of order n . Its elements $I, \sigma, \sigma^2, \dots, \sigma^{n-1}$ are linearly independent over F . (Theorem 14.1.) Let ξ be a primitive n th root of unity in F . Then $\xi \neq 0$ and $I + \xi^{-1}\sigma + \xi^{-2}\sigma^2 + \dots + \xi^{-n+1}\sigma^{n-1}$ is a non-zero endomorphism of E as a vector space over F . Thus for some $a \in E$, $c = I(a) + \xi^{-1}\sigma(a) + \xi^{-2}\sigma^2(a) + \dots + \xi^{-n+1}\sigma^{n-1}(a) \neq 0$.

$$\begin{aligned} \text{Now } \sigma(c) &= \sigma(a) + \xi^{-1}\sigma^2(a) + \xi^{-2}\sigma^3(a) + \dots + \xi^{-n+1}\sigma^n(a) + I(a) \\ &= \xi[\xi^{-1}\sigma(a) + \xi^{-2}\sigma^2(a) + \dots + \xi^{-n+1}\sigma^{n-1}(a) + I(a)] \\ &= \xi c, \text{ since } \sigma^n = I, \xi^n = 1. \end{aligned}$$

In general $\sigma^k(c) = \xi^k c \forall k = 1, 2, \dots, n-1$. So c^n is in the fixed field under $G(E, F)$. Hence $\sigma^k(c^n) = [\sigma^k(c)]^n = \xi^{kn}c^n = c^n$ for $k = 1, 2, \dots, n-1$. So $c^n \in F$. Put $b = c^n$. Thus $x^n - b$ is a polynomial over F whose roots are $c, \xi c, \xi^2 c, \dots, \xi^{n-1}c$. However $\sigma^k(c) = \xi^k c$ shows that $c, \xi c, \xi^2 c, \dots, \xi^{n-1}c$ are conjugates over F . Thus the minimal polynomial $f(x)$ of c over F is at least of degree n . As $f(x) \mid (x^n - b)$ we get $f(x) = x^n - b$, so it is itself the minimal polynomial of c . Hence $x^n - b$ is irreducible over F . Now $[F(c) : F] = n$, $[E : F] = n$, $F(c) \subseteq E$ gives $E = F(c)$. So E is the splitting field of $x^n - b$ over F and $x^n - b$ is an irreducible polynomial over F . This proves the theorem. ■

Definition 14.25. Let $p(x)$ be any (non-zero) polynomial over a field F and E be the splitting field of $p(x)$ over F . The group of all F -automorphisms of E is called the *Galois group* of $p(x)$ over F .

Now we give a relationship between the nature of $p(x)$ and a Galois group. Hence forth we shall confine ourselves to the case of a field F which has all the primitive n th roots of unity for every n , unless otherwise mentioned.

Theorem 14.26. Let F be a field of characteristic zero containing all roots of unity then a polynomial $p(x)$ over F is solvable by radicals if and only if its Galois group over F is a solvable group.

Proof: Let $p(x)$ be solvable by radicals. That means the splitting field E of $p(x)$ over F is contained in a radical extension of F . Hence we can find a sequence of fields

$$F_0 = F \subseteq F_1 = F_0(\omega_1) \subseteq F_2 = F_1(\omega_2) \subseteq F_3 = F_2(\omega_3) \subseteq \dots \subseteq F_k = F_{k-1}(\omega_k)$$

such that $\omega_1^{r_1} \in F_0, \omega_1^{r_2} \in F_1, \dots, \omega_k^{r_k} \in F_{k-1}$ for integers $r_i \geq 1$

($j=1, \dots, k$) and $E \subseteq F_k$. In view of Theorem 14.23, we can suppose that F_k is a normal extension of F . Consider the groups $G_i = G(F_k, F_i) \forall i=0, 1, 2, \dots, k-1$.

For any $i=1, 2, \dots, k$, $F_i = F_{i-1}(\omega_i)$ with $\omega_i^{r_i} \in F_{i-1}$ gives that ω_i is a root of the polynomial $x^{r_i} - \omega_i^{r_i}$. Let ζ_i be a primitive r_i th root of unity in F . Then $\zeta_i \in F_{i-1}$ and the r_i roots $\omega_i, \zeta_i \omega_i, \dots, \zeta_i^{r_i-1} \omega_i$ of $x^{r_i} - \omega_i^{r_i}$ are all in F_i . So that F_i is a splitting field of $x^{r_i} - \omega_i^{r_i}$ over F_{i-1} . By Theorem 14.21, $G(F_i, F_{i-1})$ is Abelian. Now F_k is also a normal extension of F_{i-1} . By the Fundamental Theorem of Galois Theory (Theorem 14.10) $G_i = G(F_k, F_i)$ is a normal subgroup of $G_{i-1} = G(F_k, F_{i-1})$ and $G(F_i, F_{i-1}) \cong \frac{G_{i-1}}{G_i}$. So in the descending chain

$G_0 = G(F_k, F_0) \supseteq G_1 \supseteq G_2 \supseteq \dots \supseteq G_{k-1} \supseteq G_k = \{I\}$, each G_i is normal subgroup of G_{i-1} and $\frac{G_{i-1}}{G_i}$ is an Abelian group. This proves that $G(F_k, F)$ is a solvable group. Since E is a normal extension of F ,

$G(F_k, E)$ is a normal subgroup of $G(F_k, F)$ and $G(E, F) = \frac{G(F_k, F)}{G(F_k, E)}$.

Since the quotient of any solvable group is solvable, $G(E, F)$ is a solvable group.

Conversely let $G(E, F)$ be a solvable group. We can find a sub-normal series

$$G_0 = G(E, F) \supseteq G_1 \supseteq G_2 \supseteq \dots \supseteq G_{k-1} \supseteq G_k = \{I\}$$

such that $\frac{G_i}{G_{i+1}}$ is cyclic group for $i=0, 1, 2, \dots, k-1$. Let F_i be

the fixed field under G_i . Then G_i is the group of all F_i -automorphisms of $F_k (= E)$ i.e., $G(E, F_i)$. As G_{i+1} is a normal subgroup of G_i , we get

F_{i+1} is a normal extension of F_i and $G(F_{i+1}, F_i) \cong \frac{G_i}{G_{i+1}}$ is cyclic group.

Now F_i contains all the primitive n th roots of unity for every positive integer n , as the same holds for F . So by Theorem 14.24, there exists a $b_i \in F_i$ and a positive integer r_i such that $x^{r_i} - b_i$ is irreducible over F_i and F_{i+1} is the splitting field of $x^{r_i} - b_i$ over F_i . Let ω_i be any root of $x^{r_i} - b_i$ in F_{i+1} . Then $F_{i+1} = F_i(\omega_i)$ and $\omega_i^{r_i} \in F_i$. Hence we have a sequence $F = F_0 \subseteq F_1 \subseteq F_2 \subseteq \dots \subseteq F_k = E$ of fields such that $F_{i+1} = F_i(\omega_i)$ for some $\omega_i \in F_{i+1}$ and $\omega_i^{r_i} \in F_i$. Hence E is an extension of F by radicals. This proves that $p(x)$ is solvable by radicals. ■

Theorem 14.27. Any quartic over F is solvable by radicals.

Proof: Let $p(x) = a_0x^4 + a_1x^3 + a_2x^2 + a_3x + a_4$, $a_0 \neq 0$ be any quartic over F . Let E be its splitting field over F . Then $E = F(\alpha_1, \alpha_2, \alpha_3, \alpha_4)$ where $\alpha_1, \alpha_2, \alpha_3, \alpha_4$ are roots of $f(x)$. Now for any $\sigma \in G(E, F)$, $\sigma(\alpha_i)$

is a root of $f(x) \forall i=1, 2, 3, 4$. Thus σ induces a permutation of $(\alpha_1, \alpha_2, \alpha_3, \alpha_4)$. Let S_4 be the symmetric group of all permutations on the set $\{\alpha_1, \alpha_2, \alpha_3, \alpha_4\}$. Define mapping $f: G(E, F) \rightarrow S_4$ by $f(\sigma) =$ the restriction of σ to $\{\alpha_1, \alpha_2, \alpha_3, \alpha_4\}$. This f is a monomorphism and so $G \cong f(G)$. Since S_4 is solvable (Example 7 Chapter 5), each subgroup of S_4 is solvable (Theorem 5.8). In particular $f(G)$ is solvable, hence G is solvable. Thus $p(x)$ is solvable over F (Theorem 14.26). ■

From Examples 7 and 8 and the above theorem it follows that all polynomials of degree 2, 3 and 4 over F are solvable by radicals.

Consider any field F and the field $F(x_1, x_2, \dots, x_n)$ in n indeterminates over F . Let S_n be the symmetric group of degree n acting on the set $\{1, 2, \dots, n\}$. For each $\sigma \in S_n$ we can define an automorphism σ' of $F(x_1, x_2, \dots, x_n)$ as follows. Consider any rational function $r(x_1, x_2, \dots, x_n)$. Then

$$\sigma'[r(x_1, x_2, \dots, x_n)] = r[x_{\sigma(1)}, x_{\sigma(2)}, \dots, x_{\sigma(n)}]$$

Thus $\sigma \rightarrow \sigma'$ is a monomorphism of S_n into the group of all F -automorphisms of $F(x_1, x_2, \dots, x_n)$. So we can treat S_n as a group of F -automorphisms of $F(x_1, x_2, \dots, x_n)$. Let S be the fixed field under S_n . Each member of S is called a *symmetric function*.

Consider $a_1 = x_1 + x_2 + \dots + x_n$.

$$a_2 = \sum x_i x_j$$

$$a_3 = \sum_{i < j < k} x_i x_j x_k$$

$$\dots\dots\dots$$

$$a_n = x_1 x_2 \dots x_n$$

For example when $n=3$, then $a_2 = x_1 + x_2 + x_3$, $a_2 = x_1 x_2 + x_1 x_3 + x_2 x_3$, $a_3 = x_1 x_2 x_3$; when $n=4$, $a_1 = x_1 + x_2 + x_3 + x_4$, $a_2 = x_1 x_2 + x_1 x_3 + x_1 x_4 + x_2 x_3 + x_2 x_4 + x_3 x_4$, $a_3 = x_1 x_2 x_3 + x_1 x_2 x_4 + x_1 x_3 x_4 + x_2 x_3 x_4$, $a_4 = x_1 x_2 x_3 x_4$.

It can be easily seen that each of $a_1, a_2, \dots, a_n$ is a symmetric function i.e., $a_1, a_2, \dots, a_n$ all belong to S . Each of $a_1, a_2, \dots, a_n$ is called an *elementary symmetric function* of $x_1, x_2, \dots, x_n$. Note that $F(a_1, a_2, \dots, a_n) \subseteq S$. We establish the following theorem.

Theorem 14.28. Let F be any field and let $F(x_1, x_2, \dots, x_n)$ be the field of rational functions in n indeterminates $x_1, x_2, \dots, x_n$. Suppose that S is the field of symmetric functions. Then

- $[F(x_1, x_2, \dots, x_n) : S] = n!$
- $S = F(a_1, a_2, \dots, a_n)$ where $a_1, a_2, \dots, a_n$ are elementary symmetric functions in $x_1, x_2, \dots, x_n$.

(iii) the Galois group $G[F(x_1, x_2, \dots, x_n): S]$ is isomorphic to S_n .

Proof: (i) As remarked before we regard S_n as a group of F automorphisms of $F(x_1, x_2, \dots, x_n)$ and S is the fixed field under S_n , so $[F(x_1, x_2, \dots, x_n): S] = o(S_n) = n!$, (Theorem 14.8).

(ii) Since $F(a_1, a_2, \dots, a_n) \subseteq S$,

$$[F(x_1, x_2, \dots, x_n) : F(a_1, a_2, \dots, a_n)] \geq [F(x_1, x_2, \dots, x_n) : S] = n!$$

Consider the polynomial $p(t) = t^n - a_1 t^{n-1} + a_2 t^{n-2} + \dots + (-1)^n a_n = (t - x_1)(t - x_2) \dots (t - x_n)$.

Each of $x_1, x_2, \dots, x_n$ is a root of $p(t)$ and these roots generate $F(x_1, x_2, \dots, x_n)$ over $F(a_1, a_2, \dots, a_n)$. Hence $F(x_1, x_2, \dots, x_n)$ is the splitting field of $p(t)$ over $F(a_1, a_2, \dots, a_n)$. However $\deg p(t) = n$. Hence $[F(x_1, x_2, \dots, x_n) : F(a_1, a_2, \dots, a_n)] \leq n!$ (Theorem 13.30). We conclude that $[F(x_1, x_2, \dots, x_n) : F(a_1, a_2, \dots, a_n)] = n!$. Thus

$$\begin{aligned} n! &= [F(x_1, x_2, \dots, x_n) : F(a_1, a_2, \dots, a_n)] \\ &= [F(x_1, x_2, \dots, x_n) : S][S : F(a_1, a_2, \dots, a_n)] \\ &= n! [S : F(a_1, a_2, \dots, a_n)]. \end{aligned}$$

So we get $[S : F(a_1, a_2, \dots, a_n)] = 1$ and

$$S = F(a_1, a_2, \dots, a_n). \text{ This proves (ii).}$$

(iii) Since $[F(x_1, x_2, \dots, x_n) : S] = n!$ we have $o\{G[F(x_1, x_2, \dots, x_n), S]\} \leq n!$. However S_n is a group of order $n!$ contained in $G[F(x_1, x_2, \dots, x_n), S]$.

Hence $G[F(x_1, x_2, \dots, x_n), S] = S_n$. This proves (iii). ■

Definition 14.29. Let $x_1, x_2, \dots, x_n$ be n indeterminates over any field F and $a_1, a_2, a_3, \dots, a_n$ be the elementary symmetric functions in $x_1, x_2, \dots, x_n$. Then the polynomial

$P_n(t) = t^n - a_1 t^{n-1} + a_2 t^{n-2} - a_3 t^{n-3} + \dots + (-1)^n a_n$ is called *n th generic polynomial*.

Theorem 14.30. (Abel). The generic polynomial $p_n(t)$ of degree $n \geq 5$ is not solvable by radicals over $F(a_1, a_2, \dots, a_n)$.

Proof: Since the Galois group of $p_n(t)$ over $F(a_1, a_2, \dots, a_n)$ is S_n , (Theorem 14.28) and S_n is not solvable for $n \geq 5$ (Corollary 5.42), we get that $p_n(t)$ is not solvable by radicals. ■

REMARK 1. We have been assuming that F is of characteristic zero and for each positive integer n , F contains all n th roots of unity. However in the case of Theorem 14.28 there is no need of imposing any restriction of F . Indeed many of the above results can be established for any field of characteristic zero.

REMARK 2. Above theorem gives that for each $n \geq 5$ there exists a polynomial of degree n over some appropriate field which is not solvable by radicals. Hence in general, polynomials of degree ≥ 5 are not solvable by radicals.

WORKED-OUT EXERCISES

Exercise 1. Prove that a symmetric polynomial in $x_1, \dots, x_n$ is a polynomial in the elementary symmetric functions in $x_1, \dots, x_n$.

Solution. In case $n=2$; $a_1=x_1+x_2$, $a_2=x_1x_2$. Let $f(x_1, x_2)$ be a symmetric polynomial in x_1 and x_2 . Rearrange $f(x_1, x_2)$ in descending powers of x_1 , yielding $f(x_1, x_2) = A_t x_1^t + A_{t-1} x_1^{t-1} + \dots + A_0$ where $A_t, A_{t-1}, \dots, A_0$ are polynomials in x_2 . If we replace x_2 by $a_1 - x_1$ in $f(x_1, x_2)$ then $f(x_1, x_2)$ becomes a polynomial in x_1 . We get that $f(x_1, x_2) = B_u x_1^u + B_{u-1} x_1^{u-1} + \dots + B_0 = g(x_1)$, say; where $B_u, B_{u-1}, \dots, B_0$ are polynomials in a_1 .

Then $g(z) = B_u z^u + B_{u-1} z^{u-1} + \dots + B_0$. Put $h(z) = z^2 - a_1 z + a_2$. By division algorithm in $F[a_1, a_2][z]$ we get that $g(z) = h(z)q(z) + Cz + D$ where $q(z) \in F[a_1, a_2][z]$ and $C, D \in F[a_1, a_2]$.

Now $g(x_1) = h(x_1)q(x_1) + Cx_1 + D \Rightarrow f(x_1, x_2) = Cx_1 + D$ as $h(x_1) = 0$ in view of the fact that $x_1^2 - (x_1 + x_2)x_1 + x_1x_2 = 0$. Interchanging x_1 and x_2 and keeping in mind that $f(x_1, x_2) = f(x_2, x_1)$ [as $f(x_1, x_2)$ is a symmetric polynomial in x_1 and x_2], we get that $f(x_1, x_2) = Cx_2 + D$; i.e., $Cx_1 + D = Cx_2 + D \Rightarrow C(x_1 - x_2) = 0 \Rightarrow C = 0$ and so $f(x_1, x_2) = D$. Thus $f(x_1, x_2) \in F[a_1, a_2]$. In other words $f(x_1, x_2)$ is a polynomial in elementary symmetric functions.

Let us suppose now $n > 2$ and the result to be true for $n-1$ indeterminates. If $t_1, t_2, \dots, t_n$ are the elementary symmetric functions of the $n-1$ indeterminates $x_2, x_3, \dots, x_n$, we have $a_1 = x_1 + t_1$; $a_2 = x_1 t_1 + t_2$, $a_3 = x_1 t_2 + t_3$ and so on till $a_{n-1} = x_1 t_{n-2} + t_{n-1}$. These formulae give rise to $t_1 = -x_1 + a_1$, $t_2 = -x_1 t_1 + a_2 = -x_1(-x_1 + a_1) + a_2 = x_1^2 - x_1 a_1 + a_2$, $t_3 = -x_1(x_1^2 - x_1 a_1 + a_2) + a_3 = -x_1^3 + x_1^2 a_1 - x_1 a_2 + a_3, \dots$, and $t_{n-1} = (-1)^{n-1} [x_1^{n-1} - x_1^{n-2} a_1 + \dots + (-1)^{n-1} a_{n-1}]$.

Let $f(x_1, x_2, \dots, x_n)$ be a symmetric polynomial in $x_1, x_2, \dots, x_n$. Arrange it in descending powers of x_1 to yield $f(x_1, x_2, \dots, x_n) = A_p x_1^p + A_{p-1} x_1^{p-1} + \dots + A_0$ where $A_0, A_1, \dots, A_p$ are symmetric polynomials in $x_2, x_3, \dots, x_n$; which by induction hypothesis can be expressed as polynomials in $t_1, t_2, \dots, t_{n-1}$.

Using relations between t 's and a 's we get that $A_p, A_{p-1}, \dots, A_0$ can be expressed as polynomials in $x_1, a_1, a_2, \dots, a_{n-1}$. (Replace $t_1, t_2, \dots, t_{n-1}$ by their values in terms of $x_1, a_1, a_2, \dots, a_{n-1}$).

Hence $f(x_1, x_2, \dots, x_n) = B_m x_1^m + B_{m-1} x_1^{m-1} + \dots + B_0 = g(x_1)$, say, where $B_m, B_{m-1}, \dots, B_0 \in F[a_1, a_2, \dots, a_{n-1}]$. Then $g(z) = B_m z^m + B_{m-1} z^{m-1} + \dots + B_0$. Let $h(z) = z^n - a_1 z^{n-1} + a_2 z^{n-2} - \dots + (-1)^n a_n$. Applying division algorithm in $F[a_1, a_2, \dots, a_n][z]$ we get that $g(z) = h(z)q(z) + C_{n-1} z^{n-1} + C_{n-2} z^{n-2} + \dots + C_0$, where $C_{n-1}, C_{n-2}, \dots, C_0 \in F[a_1, a_2, \dots, a_n]$.

$$= x^n - (\sigma_{1,n-1} + x_n)x^{n-1} + \sum_{i=2}^{n-1} (-1)^i (\sigma_{i,n-1} + \sigma_{i-1,n-1}x_n)x^{n-i} \\ + (-1)^n \sigma_{n-1,n-1}x_n.$$

Now $\sigma_{1,n-1} + x_n = x_1 + x_2 + \dots + x_{n-1} + x_n = \sigma_{1,n}$; $\sigma_{i,n-1} + \sigma_{i-1,n-1}x_n = \sigma_{i,n}$ as now we have the sum of all the products of n indeterminates $x_1, x_2, \dots, x_n$, taken i at a time and finally $\sigma_{n-1,n-1}x_n = x_1x_2 \dots x_{n-1}x_n = \sigma_{n,n}$.

$$\text{Thus } \prod_{i=1}^n (x - x_i) = x^n - \sigma_{1,n}x^{n-1} + \sum_{i=2}^{n-1} (-1)^i \sigma_{i,n}x^{n-i} + (-1)^n \sigma_{n,n} \\ = x^n + \sum_{i=1}^n (-1)^i \sigma_{i,n} x^{n-i}.$$

PROBLEMS

1. Prove that the elementary symmetric functions in $x_1, x_2, \dots, x_n$ are indeed symmetric functions in $x_1, \dots, x_n$.

2. Prove Newton's identities, namely if $\alpha_1, \alpha_2, \dots, \alpha_n$ are the roots of $f(x) = x^n + a_1x^{n-1} + a_2x^{n-2} + \dots + a_n$ and if $s_k = \alpha_1^k + \alpha_2^k + \dots + \alpha_n^k$ then:

$$(a) s_k + a_1s_{k-1} + a_2s_{k-2} + \dots + a_{k-1}s_1 + ka_k = 0 \text{ for all } k=1, 2, \dots, n.$$

$$(b) s_k + a_1s_{k-1} + \dots + a_ns_{k-n} = 0 \text{ for all } k > n.$$

3. Express the following as polynomials in the elementary symmetric functions in x_1, x_2, x_3 .

$$(a) x_1^2 + x_2^2 + x_3^2; \quad (b) x_1^3 + x_2^3 + x_3^3;$$

$$(c) (x_1 - x_2)^2(x_1 - x_3)^2(x_2 - x_3)^2; \quad (d) -(4q^3 + 27r^2) \text{ where}$$

$$[\text{Ans. } (a) a_1^2 - 2a_2; \quad (b) a_1^3 - 3a_1a_2 + 3a_3; \quad (c) -(4q^3 + 27r^2) \text{ where} \\ q = a_1 - \frac{a_1^2}{3} \text{ and } r = -a_3 + \frac{a_1a_2}{3} - \frac{2a_1^3}{27}.$$

4. Let K be a normal radical extension of F of characteristic zero. Prove the following:

(i) There exist elements $a_1, a_2, \dots, a_n \in K$ and positive integers $s_1, s_2, \dots, s_n$ such that $K = F(a_1, a_2, \dots, a_n)$ and for any i , $a_i^{s_i} \in F(a_1, a_2, \dots, a_{i-1})$ for $i=2, 3, \dots, n$ and $a_1^{s_1} \in F$.

(ii) Let $s = s_1s_2 \dots s_n$ and let N be the splitting field of $x^s - 1$ over K . Prove that N is also a normal extension of F of finite degree.

(iii) N contains a primitive s th root of unity say, ζ .

(iv) If $L = F(\zeta)$ then L is a splitting field of $x^s - 1$ over F .

(v) $G(L, F)$ is Abelian.

(vi) Define $L_0 = L$, $L_i = L(a_1, a_2, \dots, a_i)$ for $i=1, 2, \dots, n$. Show

that L_i is a normal extension of L_{i-1} such that $G(L_i, L_{i-1})$ is Abelian. Deduce that $G(N, L)$ is a solvable group.

(vii) $G(N, F)$ is a solvable group.

(viii) $G(K, F)$ is a solvable group.

[Hint. (i) follows from definition. (ii) N is the splitting field of a polynomial $(x^n - 1)g(x)$, where $g(x)$ is such that K is splitting field of $g(x)$ over F . (v) follows from Theorem 14.20. (vi) follows from Lemma 14.21. (vii) follows from (v) and (vi). (iii) follows from the fact that $G(K, F) \cong \frac{G(N, F)}{G(N, K)}$.]

5. Using problem 4, show that any polynomial $f(x)$ over a field F of characteristic zero is solvable by radicals if and only if the Galois group of $f(x)$ over F is solvable. (This generalizes Theorem 14.26).

6. Prove that $x^5 - 4x + 2$ is not solvable by radicals over $\mathbb{Q}$.

4. Constructions by Ruler and Compass

There are many problems in Euclidean Geometry which had baffled Mathematicians since antiquity, but were finally solved by the application of theory of fields. In this section, we shall apply the theory of fields developed before, to Geometry and finally solve the following old problems.

1. To construct by ruler and compass a square having the same area as a circle.
2. To construct by ruler and compass a cube of twice the volume of a given cube.
3. To trisect a given angle by ruler and compass.

Before we take up these problems, let us understand what do we mean by construction with ruler and compass. For our convenience by a line, we shall always mean a straight line. In Euclidean Geometry, a ruler is only used to draw a (straight) line through two given points and a compass, together with the help of a ruler, is used to draw a circle with centre at a given point and radius equal to the distance between two given points.

Let S be a set of points in a plane, such that S contains at least two members. A line is said to be *constructible* from S if it passes through two distinct points of S . A circle is said to be *constructible* from S if its centre is a point of S and its radius is equal to the distance between two points of S . Any point of the plane which is an intersection point of either a pair of distinct lines both constructible from S or of two distinct circles both constructible from S or of a line and a circle both constructible from S , is called *constructible* from S .

ible by ruler and compass or simply constructible from S . Let $s(S)$ be the set of all points constructible from S . Consider any point P of S . Since S has at least two members there exists a point Q in S , different from P . Then the circle with centre at Q and radius PQ and the line through P and Q both are constructible from S , further P is one of their points of intersection. Consequently P is constructible by ruler and compass from S or simply constructible from S . This yields $P \in s(S)$. Hence $S \subseteq s(S)$.

If we fix up a system of co-ordinates in a plane, then we know that each point of the plane has unique pair of co-ordinates and there exists a 1-1 correspondence between the points and the ordered pairs (a, b) (co-ordinates of the points) of real numbers. Here we identify each point of the plane with its co-ordinates. Consequently we identify the plane with $\mathbf{R} \times \mathbf{R}$, where $\mathbf{R}$ is the set of all real numbers. The system of co-ordinates taken here is orthogonal with $X'OX$ as the x -axis and $Y'OY$ as the y -axis. Consider the set E consisting of only two points $O(0, 0)$ and $I(1, 0)$. Here O is called the *origin* and I is called the *unit point*. Let $E_1 = s(E)$. In the previous paragraph by taking $S = E$, we have $E \subseteq E_1$. Put $E_0 = E$. Now notice the following:

- (i) Since $X'OX$ is the line passing through two points O and I of E , $X'OX$ is constructible from E .
- (ii) Since $OI = 1$, the circle with centre at O and radius 1 is constructible from E . (This circle is called the *unit-circle*). The unit circle also meets $X'OX$ at $I'(-1, 0)$. So that I' is also constructible and $I' \in E_1$.
- (iii) Since the circle with centre I and radius 1 is constructible from E and it meets x -axis also at $A(2, 0)$; A is constructible from E . Consequently $A \in E_1$.

We leave it to the readers to find out more points constructible from E . Inductively for every $n \geq 0$, define $E_{n+1} = s(E_n)$. Let $H = \bigcup_n E_n$. Any point of the plane is said to be *constructible by ruler and compass* (or simply *constructible*) if and only if it belongs to H . Here notice that $E_n \subseteq E_{n+1}$.

Lemma 14.29. $s(H) = H$ i.e. every point constructible from H is in H (i.e. every point constructible from H is constructible).

Proof: Consider a line l constructible from H . Then there are two distinct points P and Q of H such that the line passing through P and Q is l . As $H = \bigcup_n E_n$, $P \in E_n$ and $Q \in E_m$ for some n, m . Either $n \leq m$ or $m \leq n$, to be definite let $n \leq m$. Then $E_n \subseteq E_m$ and $P, Q \in E_m$.

Consequently the line l is constructible from E_m . Consider a circle C constructible from H . There exists three points A , B and D such that A is the centre and BD is the radius of C . Again for some positive integers r, s, t , $A \in E_r$, $B \in E_s$ and $D \in E_t$. Among r, s, t , one of them, say, r is the largest. Then A, B, D are all in E_r . Consequently the circle C is constructible from E_r . Thus any line or a circle constructible from H is constructible from some E_n .

Let P be a point constructible from H . By definition P is a point of intersection of G_1 and G_2 , where each of G_i is either a line or a circle constructible from H . By what has been done in the previous paragraph, for some n, m , G_1 is constructible from E_n and G_2 is constructible from E_m . Since either $E_n \subseteq E_m$ or $E_m \subseteq E_n$, we can suppose for the sake of definiteness that $E_n \subseteq E_m$. Then G_1 and G_2 are both constructible from E_m . This yields P is constructible from E_m and hence $P \in s(E_m) = E_{m+1} \subseteq H$. Hence $s(H) \subseteq H$. However $H \subseteq s(H)$. We get $H = s(H)$. ■

A line or a circle is said to be *constructible* if it is constructible from H . Let us make the following observations:

1. If a constructible point P lies upon a constructible line l , then the line through P perpendicular to l is also constructible (Problem 1).

2. If a constructible point P does not lie upon a constructible line l , then the line through P parallel to l and the line through P perpendicular to l , both are constructible (Problems 2 and 3). To illustrate this, we notice that the points $I(1, 0)$ and $I'(-1, 0)$ are constructible points. Then the two circles with centres I and I' and having same radius II' are both constructible. Consequently the line through their point of intersection is constructible. This line is nothing else but the y -axis. Thus y -axis is constructible. The y -axis is the line through the constructible point O , perpendicular to the constructible line $X'OX$. We have seen before that point $A(2, 0)$ is constructible. The line passing through the points of intersections of two circles with their centres at O and A and both having radius 2, is constructible and this line is parallel to the constructible line $Y'OY$ and is passing through the constructible point I .

A real number α is said to be *constructible* if and only if the point $(\alpha, 0)$ is constructible. Let T be the set of all constructible real numbers. Since the points, $O(0, 0)$ and $I(1, 0)$ are both constructible we see that $O, I \in T$. Further it can be seen easily that if two real numbers α and β are constructible then $\alpha - \beta$ and $\alpha\beta^{-1}$ (if $\beta \neq 0$) are constructible (Problems 5, 7 and 9). Thus T is a subfield of $\mathbb{R}$, the field of real numbers. Further if any non-negative real number α is

constructible then $\pm\sqrt{\alpha}$ are also constructible (Problem 11). Since we do not intend to draw any diagrams we advise the readers to visualize the constructions of different real numbers mentioned above, by drawing the diagrams themselves.

Lemma 14.30. For any real number α , the point $(\alpha, 0)$ is constructible if and only if $(0, \alpha)$ is constructible.

Proof: Suppose that the point $(\alpha, 0)$ is constructible. Since the origin O is constructible, the circle with centre at the origin passing through $(\alpha, 0)$ is constructible. This circle meets y -axis at the point $(0, \alpha)$. Since y -axis is constructible, we get the point $(0, \alpha)$ is constructible.

Conversely if $(0, \alpha)$ is constructible, the circle with centre O and passing through $(0, \alpha)$ meets x -axis at $(\alpha, 0)$. Hence the lemma follows. ■

Lemma 14.31. A point (α, β) is constructible if and only if the real numbers α and β are constructible.

Proof: Let the real numbers α and β be constructible. The definition of constructible real numbers and the Lemma 14.30 yield that the points $A(\alpha, 0)$ and $B(0, \beta)$ are both constructible. Consequently the line through B parallel to x -axis and the line through A parallel to y -axis both are constructible (Problem 2). These lines intersect at the point (α, β) . Hence (α, β) is constructible.

Conversely suppose that the point $P(\alpha, \beta)$ is constructible. Then the line through P parallel to y axis meets x -axis at $(\alpha, 0)$. So the point $(\alpha, 0)$ and hence the real number α are constructible. Again the line through P parallel to x axis meets y -axis at $(0, \beta)$. So the point $(0, \beta)$ is constructible. By Lemma 14.30 the point $(\beta, 0)$ is constructible. Hence by definition β is also constructible. ■

A complex number $z = \alpha + i\beta$ where α and β are real numbers and $i = \sqrt{-1}$ is said to be constructible in case the point (α, β) is constructible. Now we prove the following:

Theorem 14.32. The set L of all constructible complex numbers is a subfield of $\mathbb{C}$, the field of complex numbers. Every root in $\mathbb{C}$ of a quadratic polynomial whose coefficients are constructible complex numbers is a constructible complex number (i.e. the root of any quadratic over L is a member of L).

Proof: Let T be the set of all constructible real numbers. As observed before T is a subfield of the field of real numbers. Let $w = \alpha + i\beta$ be a complex number. By definition w is constructible if and only if the point (α, β) is constructible. But by Lemma 14.31, the point (α, β) is constructible if and only if α, β are constructible.

Hence w is constructible if and only if $w \in T + Ti = T(i)$. Hence $L = T(i)$ is a subfield of the field of all complex numbers.

Suppose a complex number t is a root of quadratic $z^2 + az + b$ over L , (without loss of generality we can choose the quadratic to be monic). Then $t^2 + at + b = 0 \Rightarrow t = -\frac{a}{2} \pm \frac{1}{2} \sqrt{a^2 - 4b}$. Since $-\frac{a}{2} \in L$,

we have $t \in L$ if and only if $\sqrt{a^2 - 4b} \in L$. Thus if we show that for any $w \in L$, $\sqrt{w} \in L$, it will follow that roots of every quadratic over L is a member of L . So consider $w = \alpha + i\beta$; $\alpha, \beta \in T$ and let

$$z^2 = w = \alpha + i\beta \quad \dots(1)$$

Case I. $\beta = 0$. If $\alpha \geq 0$, then $z = \pm \sqrt{\alpha}$ is real. Since a square root of non-negative numbers $\in T$ is constructible, we have $z \in T \subseteq L$. If $\alpha < 0$ then $-\alpha > 0$ and $-\alpha \in T$ also gives $\pm \sqrt{-\alpha} \in T$. However $z = \pm i\sqrt{-\alpha}$. This again yields $z \in T(i) = L$.

Case II: $\beta \neq 0$. Let $z = x + iy$; x, y real, then (I) yields

$$x^2 - y^2 = \alpha \quad \dots(2)$$

$$2xy = \beta \quad \dots(3)$$

(3) yield $x \neq 0$ and $y = \frac{\beta}{2x}$.

$$\text{So by (2)} \quad x^2 - \frac{\beta^2}{4x^2} = \alpha \Rightarrow x^4 - \alpha x^2 - \frac{\beta^2}{4} = 0$$

$$\Rightarrow \left(x^2 - \frac{\alpha}{2}\right)^2 = \frac{1}{4}(\alpha^2 + \beta^2) \quad \dots(4)$$

Since $\frac{1}{4}(\alpha^2 + \beta^2) > 0$ and is in T , we have

$$\sqrt{\frac{1}{4}(\alpha^2 + \beta^2)} \in T,$$

$$\text{so by (4)} \quad x^2 - \frac{\alpha}{2} = \pm \sqrt{\frac{1}{4}(\alpha^2 + \beta^2)} \in T,$$

$\Rightarrow x^2$ is constructible, as $\frac{\alpha}{2}$ is constructible. Then, case I yields x is constructible. This implies $x \in T$ and so $y = \frac{\beta}{2x} \in T$. Hence $z = x + iy \in T(i) = L$. ■

Lemma 14.33. Let F be a subfield of the field of real numbers, given G_1 and G_2 (distinct from G_1), where each G_i is either a line or a circle whose equation has all its coefficients in field F and if (x, y) is a point of their intersection, then either both x and y are in F or are in a quadratic extension of F .

Proof: Case I: G_1 and G_2 both are lines. Let their equations be

$$ax+by+c=0 \quad \dots(1)$$

$$\text{and} \quad a'x+b'y+c'=0 \quad \dots(2)$$

with their coefficients in F . These lines can intersect only if they are not parallel. Then their points of intersection is (x_1, y_1) given by

$$x_1 = \frac{bc' - b'c}{ab' - a'b}, \quad y_1 = \frac{a'c - ac'}{ab' - a'b}$$

Clearly $x_1, y_1 \in F$.

Case II. One of G_1 and G_2 is a line and other is a circle. To be definite let G_1 be a line and G_2 be a circle. Let their respective equations be

$$ax+by+c=0 \quad \dots(3)$$

and $x^2+y^2+ex+fy+g=0$ $\dots(4)$
with coefficients in F . In (3) at least one of a and b , say a , is non-zero. Eliminating x between (3) and (4) we get

$$y^2 \left(\frac{b^2}{a^2} + 1 \right) + y \left(\frac{2bc}{a^2} - \frac{be}{a} + f \right) - \frac{ec}{a} + \frac{c^2}{a^2} + g = 0 \quad \dots(5)$$

i.e.

$$\alpha y^2 + \beta y + \gamma = 0$$

where

$$\alpha = 1 + \frac{b^2}{a^2}, \quad \beta = \frac{2bc}{a^2} - \frac{be}{a} + f, \quad \gamma = \frac{c^2}{a^2} + g - \frac{ec}{a}.$$

(5) yields

$$y = -\frac{\beta}{2\alpha} \pm \frac{1}{2\alpha} \sqrt{\beta^2 - 4\alpha\gamma}.$$

So we put $d = \sqrt{\beta^2 - 4\alpha\gamma}$, we have $d^2 \in F$. Thus $[F(d) : F] \leq 2$. Clearly $y_1, y_2 \in F(d)$, where (x_1, y_1) and (x_2, y_2) are the coordinates of points of intersections of G_1 and G_2 . (3) yields $x_1 = -\frac{c}{a} - \frac{b}{a} y_1$

and $x_2 = -\frac{c}{a} - \frac{b}{a} y_2 \in F(d)$.

Case III. G_1 and G_2 both are circles. Let their respective equations be

$$x^2+y^2+ex+fy+g=0 \quad \dots(6)$$

and $x^2+y^2+e'x+f'y+g'=0$ $\dots(7)$
with coefficients e, f, g, e', f' and g' in F . The line passing through their points of intersection is

$$(e-e')x + (f-f')y + g-g' = 0$$

The points of intersection of (6) and (7) are same as the points of intersection of G_1 and the line given by (8). So this case reduces to case II.

Hence the lemma follows. ■

Theorem 14.34. A complex number u is constructible if and only if

there exists a sequence $\{F_j\}$, $0 \leq j \leq n$ of subfields of $\mathbb{C}$, such that

$$\mathbb{Q} = F_0 \subseteq F_1 \subseteq F_2 \subseteq \dots \subseteq F_n \quad \dots(1)$$

$$[F_j : F_{j-1}] \leq 2 \text{ for all } j, 1 \leq j \leq n \quad \dots(2)$$

$$\text{and } u \in F_n. \quad \dots(3)$$

Proof: Note that a sequence $\{F_j\}$ satisfying (1), (2) and (3) is called *admissible sequence* for u . (2) $\Rightarrow F_j = F_{j-1}(c_j)$ for some $c_j \in F_j$.

Necessity: We show firstly that if two complex numbers u and v have admissible sequences, then they have a common admissible sequence. Let $\{F_j\}$, $0 \leq j \leq n$ and $\{K_j\}$, $0 \leq j \leq m$ be their respective admissible sequences. Since $[K_j : K_{j-1}] \leq 2$, $K_j = K_{j-1}(\alpha_j)$ for some $\alpha_j \in K_j$. This gives $K_1 = K_0(\alpha_1) = \mathbb{Q}(\alpha_1)$; $K_2 = K_1(\alpha_2) = \mathbb{Q}(\alpha_1, \alpha_2)$, ..., $K_j = \mathbb{Q}(\alpha_1, \alpha_2, \dots, \alpha_j)$ for $1 \leq j \leq m$. Define a sequence $\{L_k\}$, $0 \leq k \leq m+n$ as under.

$L_i = F_i$, $0 \leq i \leq n$ and $L_{n+t} = F_n(\alpha_1, \alpha_2, \dots, \alpha_t)$ for $1 \leq t \leq m$. This implies $L_{n+t} = L_{n+t-1}(\alpha_t)$ for $1 \leq t \leq m$.

Now $[L_i : L_{i-1}] = [F_i : F_{i-1}] \leq 2$ for $1 \leq i \leq n$. Since $K_{t-1} \subseteq L_{n+t-1}$ and degree of α_t over K_{t-1} is at most two, we have the degree of α_t over L_{n+t-1} is at the most two. Consequently $[L_{n+t} : L_{n+t-1}] \leq 2$. Further $K_m \subseteq L_{n+m}$ gives $v \in L_{n+m}$ and also $P_n \subseteq L_{n+m}$ gives $u \in L_{n+m}$. Hence $\{L_k\}$, $0 \leq k \leq m+n$ is an admissible sequence of u and v both.

By induction it follows that given any finite number of complex numbers each admitting an admissible sequence of subfields then they have a common admissible sequence of subfields. For each non-negative integer n , let D_n be the set of all real numbers x such that either (x, y) or (y, x) belongs to E_n for some real number y . Further let $S = \{n \mid n \text{ is a natural number and there exists admissible sequence of subfields for each member of } D_n\}$. Since $E_0 = \{(0, 0), (1, 0)\}$; $D_0 = \{0, 1\}$. Since, $0, 1 \in \mathbb{Q}$; they have trivial admissible sequences. So that $0 \in S$. Suppose that $n \in S$. Consider any $x \in D_{n+1}$. By definition $(x, y) \in E_{n+1}$ or $(y, x) \in E_{n+1}$ for some real number y . Let $(x, y) \in E_{n+1}$. Now (x, y) is a point of intersection of G_1 and G_2 where each of G_i is either a line or a circle constructible from E_n . Since a line constructible from E_n is determined by two points of E_n and a circle constructible from E_n is determined by three points (the centre, the two points distance between which, is the radius); G_1 and G_2 both are determined by finitely many points of E_n . If F is the set of co-ordinates of these finitely many points, clearly F is a finite subset of D_n . By induction hypothesis each member of F has an admissible sequence of subfields. Consequently these real numbers have a common admissible sequence, say $\{K_j\}$, $0 \leq j \leq m$. Since all the members

of F are in K_m , the equations of G_1 and G_2 are over K_m . By Lemma 14.33, both x and y are in a field extension L of K_m of degree at most 2 over K_m . If we put $K_{m+1} = L$, we see that $x, y \in K_{m+1}$ and $\{K_j\}$, $0 \leq j \leq m+1$ is their admissible sequence. Similar considerations hold in case $(y, x) \in E_{n+1}$. Hence every member x of D_{n+1} has an admissible sequence of subfields. So that $n+1 \in S$. Hence by induction, for every $n \geq 0$, every member of D_n has an admissible sequence.

Let $u = x + iy$ be a constructible complex number. Then for some n , $(x, y) \in E_n$. As proved above both x and y have admissible sequences of subfields. So they have a common admissible sequence $\{F_j\}$, $0 \leq j \leq q$. We put $F_{q+1} = F_q(i)$. Then $[F_{q+1} : F_q] = 2$, $x, y \in F_q \subseteq F_{q+1}$. Hence $\{F_j\}$, $0 \leq j \leq q+1$ is an admissible sequence of subfields for u .

Sufficiency. Suppose $u = x + iy$ has an admissible sequence of subfields $\{F_j\}$, $0 \leq j \leq q$. Let L be the field of all constructible complex numbers. We prove by induction on j that $F_j \subseteq L$. Clearly $F_0 = \mathbb{Q} \subseteq L$. Suppose for some j , $F_j \subseteq L$. Now $F_{j+1} = F_j(a_{j+1})$ for some $a_{j+1} \in F_{j+1}$. If $F_j = F_{j+1}$ trivially $F_{j+1} \subseteq L$. If $F_j \neq F_{j+1}$ then a_{j+1} is quadratic over F_j and hence over L . However every root of a quadratic over L is in L . So $a_{j+1} \in L$. This yields that $F_{j+1} \subseteq L$. Hence by induction $F_j \subseteq L$ for all j . In particular $F_q \subseteq L$. Then $u \in F_q$ implies $u \in L$ and so u is constructible. ■

Corollary 14.35. Any constructible complex number is algebraic over $\mathbb{Q}$ of degree, a power of 2.

Proof: Let u be a constructible complex number. By the above theorem u admits an admissible sequence of subfields

$$\mathbb{Q} = K_0 \subseteq K_1 \subseteq K_2 \subseteq \dots \subseteq K_m.$$

Here $u \in K_m$. Now for every j , with $1 \leq j \leq m$, $[K_j : K_{j-1}] = 1$ or 2 . From $[K_m : \mathbb{Q}] = [K_m : K_{m-1}][K_{m-1} : K_{m-2}] \dots [K_1 : K_0]$, we get that $[K_m : \mathbb{Q}] = 2^r$ for some $r \leq m$.

Now $u \in \mathbb{Q}_m$ implies $\mathbb{Q}(u) \subseteq \mathbb{Q}_m$. So $2^r = [\mathbb{Q}_m : \mathbb{Q}] = [\mathbb{Q}_m : \mathbb{Q}(u)][\mathbb{Q}(u) : \mathbb{Q}]$ yields that $[\mathbb{Q}(u) : \mathbb{Q}]$ is a power of 2. Hence the degree of u over $\mathbb{Q}$ is a power of 2. ■

The following example shows that even if a complex number is algebraic over $\mathbb{Q}$ of degree, a power of 2, it need not be constructible. By algebraic over $\mathbb{Q}$ of degree, a power of 2, it need not be constructible. By

EXAMPLE 9. Consider the polynomial $f(x) = x^3 - 4x - 2$. By Eisenstein's Criterion $f(x)$ is irreducible over $\mathbb{Q}$. By Ferrari's Method. its resolving cubic is $g(y) = y^3 - 2y - 2$.

If l is a root of $g(y)$, then

$$x_1 = \frac{\sqrt{2l} + \sqrt{4\sqrt{l^2 - 2} - 2l}}{2}$$

$$x_2 = \frac{\sqrt{2l} - \sqrt{4\sqrt{l^2 - 2} - 2i}}{2}$$

are two roots of $f(x)$. Now $g(y)$ is also irreducible, so the degree of l over $\mathbf{Q}$ is 3. Hence l is not constructible (Corollary 14.35). If both x_1 and x_2 were constructible then $x_1 + x_2$ would be constructible. This would imply $\sqrt{2l}$ is constructible, which in turn would give $l = 1/2 \cdot (\sqrt{2l})^2$ is constructible. This a contradiction. Hence at least one of x_1 and x_2 is not constructible, while degree of both x_1 and x_2 over $\mathbf{Q}$ is $4 = 2^2$, a power of 2.

A line segment is said to be *constructible* if its end-points are constructible. A polygon is said to be *constructible* if each of its sides is constructible. An angle is said to be *constructible* if its vertex and arms are constructible.

Lemma 14.36. *If a line segment is constructible then its length is a constructible number.*

Proof: Consider a constructible line segment with $A(a_1, b_1)$ and $B(a_2, b_2)$ as its end-points. By definition A and B are constructible points. Therefore a_1, a_2, b_1, b_2 are all constructible numbers. Since the set of all constructible real numbers is a field and the square root of every non-negative constructible real number is constructible, $AB = \sqrt{(a_1 - a_2)^2 + (b_1 - b_2)^2}$ is constructible. Hence the lemma follows. ■

Above lemma shows that if for a given line segment we can show that its length is not a constructible number, it follows that the line segment is not constructible.

Now we are prepared to solve the problems stated at the start of this section.

Lemma 14.37. *The polynomial $8x^3 - 6x - 1$ is irreducible over rational numbers.*

Proof. Let $f(x) = 8x^3 - 6x - 1$. It is enough to show that $f(x)$ is irreducible over $\mathbf{Z}$ (Why?).

$$\text{Now } f(x-1) = 8x^3 - 24x^2 + 18x - 3 \quad \dots(1)$$

$f(x)$ is irreducible if and only if $f(x-1)$ is irreducible.

Now 3 is a prime number which does not divide the leading coefficient of $f(x-1)$, divides all the other coefficients but $9 = 3^2$ does not divide the constant term. Hence by Eisenstein's Criterion $f(x-1)$ is irreducible. Hence the lemma follows. ■

Theorem 14.38. *In general an angle cannot be trisected by ruler and compass.*

Proof: The result will follow if we can show that there exists a constructible angle which cannot be trisected. The point $A(1, \sqrt{3})$ is

constructible. (Draw a diagram). $\angle AOX = \pi/3$ radians. So that $\angle AOX$ is constructible. If this angle can be trisected by ruler and compass, a line OC can be drawn such that $\angle COX = \pi/9$. Let this line meet the unit circle in D . Then the point D and the line segment DE perpendicular to OX (E on x -axis) are constructible. Consequently by Lemma 14.36, $OE = \cos \pi/9$ is constructible. Put $x = \cos \pi/9$. Putting $\theta = \pi/9$, in $\cos 3\theta = 4 \cos^3 \theta - 3 \cos \theta$, we get that $4x^3 - 3x - \frac{1}{2} = 0$ i.e., $8x^3 - 6x - 1 = 0$.

Since by Lemma 14.37, $8x^3 - 6x - 1$ is irreducible over $\mathbb{Q}$, we get that x is of degree 3 over $\mathbb{Q}$. But 3 is not a power of 2. Hence x is not constructible (Corollary 14.35). This is a contradiction. Hence the result follows ■

It should be emphasized that there are angles e.g. $\frac{\pi}{2}$ radians, which can be trisected by ruler and compass.

A cube each of whose sides is one unit is called a unit cube. Theorem 14.39. A cube of twice the volume of the unit cube is not constructible by ruler and compass.

Proof: The volume of unit cube is 1, so if we can construct by ruler and compass a cube of double the volume of unit cube, then its volume is 2 and hence its one side is of length $\sqrt[3]{2}$, is, constructible. Then (by Corollary 14.35) $\sqrt[3]{2}$ is of degree, a power of 2, over $\mathbb{Q}$. However the minimal polynomial of $\sqrt[3]{2}$ over $\mathbb{Q}$ is $x^3 - 2$, which is of degree not a power of 2. This is a contradiction. Hence the result follows. ■

Theorem 14.40. A square of the area equal to the area of the unit circle is not constructible by ruler and compass.

Proof: The area of a unit circle is π . So if a square of area π is constructible, it has each side of length $\sqrt{\pi}$. So $\sqrt{\pi}$ is a constructible number. Consequently $\pi = (\sqrt{\pi})^2$ is a constructible number and hence π is algebraic over $\mathbb{Q}$ (Corollary 14.35). This contradicts the fact that π is a transcendental number. Hence the result follows. ■

PROBLEMS

The students are advised to give analytic proofs of the problems 1 to 11.

1. If a constructible point P lies upon a constructible line l , show that the line through P perpendicular to l is also constructible.

[Hint. There exists a constructible point Q on l different from P . The circle with centre P and radius PQ meets l at another point say T . The circle with centres at Q and T and of same radius QT

are constructible. The line of their intersection is the desired line through P perpendicular to l .]

2. If P is a constructible point which does not lie upon a constructible line l , show that the line through P perpendicular to l is constructible.

[Hint. Choose a constructible point Q on l . The circle with centre P and radius PQ meets l in general at two points, one of them is Q and other is, say T . If T and Q coincide then the line through P and Q is the desired line. If T and Q are distinct, then the two circles with centres Q and T and having the same radius QT are constructible. The line of intersection of these circles is the desired line.]

3. If a constructible point P does not lie upon a constructible line l , show that the line through P , parallel to l is constructible.

[Hint. By problem 2, the line l' through P perpendicular to l is constructible. By problem 1, the line m through P perpendicular to l' is constructible. The line m is the desired line.]

4. If a real number α is constructible, show that $-\alpha$ is constructible.

[Hint. The point $(\alpha, 0)$ is constructible. The circle with centre at the origin and passing through $(\alpha, 0)$ is constructible. This circle also meets x -axis at $(-\alpha, 0)$.]

5. If two real numbers α and β are constructible, show that the real numbers $\alpha \pm \beta$ are constructible.

[Hint. The circle with centre $A(\alpha, 0)$ and radius OB where $B(\beta, 0)$ is constructible. This circle meets x -axis at the points $(\alpha + \beta, 0)$ and $(\alpha - \beta, 0)$.]

6. If α and β are two positive constructible numbers show that $\alpha\beta$ is constructible.

[Hint. The line through $A(0, 1)$ and $B(\alpha, 0)$ is constructible. The line through $E(1, \beta)$ parallel to AB is constructible and it meets x -axis at the point $(\alpha\beta, 0)$.]

7. Using problems 4 and 6 prove that the product of any two constructible real numbers is constructible.

8. If a positive number β is constructible show that $1/\beta$ is constructible.

[Hint. The line through $(1, 0)$ parallel to the line joining the points $(\beta, 0)$ and $(0, 1)$ is constructible and it meets y -axis at $(0, 1/\beta)$.]

9. Using problems 4 and 8, prove that for any constructible non-zero real number α , $1/\alpha$ is constructible.

10. Deduce from problems 5, 7 and 9 that the set of constructible real numbers is a field under usual addition and multiplication

of real numbers.

11. For any constructible positive number α , show that $\pm\sqrt{\alpha}$ are constructible.

[Hint. The circle with diameter joining the points $(-\alpha, 0)$ and $(1, 0)$ is constructible and the same meets y -axis at $(0, \sqrt{\alpha})$ and $(0, -\sqrt{\alpha})$. Hence $\pm\sqrt{\alpha}$ are constructible.]

12. Prove that an angle of α radians is constructible if and only if the real number $\cos \alpha$ ($\sin \alpha$) is constructible.

13. Prove that an angle of α radians is constructible if and only if the point $(\cos \alpha, \sin \alpha)$ is constructible.

14. Show that the set A of all real numbers α , such that $\cos \alpha$ is constructible is a subgroup of the additive group of real numbers.

Further for any $\alpha \in A$, $\frac{\alpha}{2} \in A$.

15. Show that $\cos \frac{\alpha}{5}$ is a constructible number and that $\frac{\pi}{5}$ is a constructible angle.

16. Given a positive integer n , show that an angle of n degrees is constructible if and only if n is a multiple of 3.

[Hint. By problem 15, angle of 36 degrees is constructible. From problem 14 it follows that an angle of degrees 9 is constructible. Let t be the least positive integer such that an angle of degrees t is constructible. Then $t \leq 9$. Consider $9 - t$ and conclude that $t=3$.]

17. Prove that a regular n -gon is constructible if and only if the angle of $\frac{\pi}{n}$ radians is constructible.

18. Using problems 16 and 17 deduce that the regular pentagon and hexagon are constructible, but a regular 9-gon is not constructible.

19. Prove that the angles of degrees 15, 6 and 3 cannot be trisected by ruler and compass.

20. Prove the following:

(i) $x^3 + x^2 - 2x - 1$ is irreducible over rationals.

(ii) $2 \cos \frac{2\pi}{7}$ satisfies the polynomial in (i).

(iii) $\frac{2\pi}{7}$ is not a constructible number.

(iv) A regular heptagon is not constructible.

21. Using the fact that for any $n \geq 1$, the cyclotomic polynomial $\phi_n(x)$ is irreducible over $\mathbb{Q}$, prove the following:

(i) The minimal polynomial of a primitive n th root of unity over $\mathbb{Q}$ is $\phi_n(x)$.

(ii) Using Theorem 14.34 deduce that a primitive n th root of unity is constructible then the degree of $\phi_n(x)$ i.e. $\phi(n)$ is equal to a power of 2, [where $\phi(n)$ is totient function.]

(iii) Using the fact that the Galois group of a primitive n th root u is commutative and the converse of Lagrange's theorem holds for Abelian groups, show that if $\phi(n)$ is power of 2 then u is constructible.

(iv) Show that a regular n -gon is constructible if and only if primitive n th root of unity is constructible.

[Hint. Use problems 12, 17 and the definition of constructible complex number.]

(v) (Gauss) A regular n -gon is constructible if and only if $\phi(n)$ is a power of 2.

[Hint. Use parts (ii), (iii) and (iv) above.]

(vi) For a prime p , show that a regular p -gon is constructible if and only if p is a Fermat's prime i.e. $p=2^{2^n}+1$ for some $n \geq 0$.

[Hint. $\phi(p)=p-1$.]

RINGS WITH CHAIN CONDITIONS

In Lemma 10.27 we proved that in a PID every ascending chain of ideals becomes stationary after a finite number of terms. In this chapter we shall undertake the study of some of the properties of rings satisfying similar conditions for one-sided ideals. We shall also give a brief account of the rings which do not admit an infinite strictly descending chain of one-sided ideals.

1. Noetherian Rings

Definition 15.1. A ring in which every strictly ascending chain of right (left) ideals is finite, is called a *Right (Left) Noetherian* ring.

Equivalently one can define a right (left) Noetherian ring to be a ring in which for every infinite ascending chain $I_1 \subseteq I_2 \subseteq I_3 \subseteq \dots$ of right (left) ideals there exists a positive integer n such that $I_m = I_n$ $\forall m \geq n$.

REMARK. A right (left) Noetherian ring is also known as a ring with ACC (i.e. Ascending Chain Condition) on right (left) ideals.

Definition 15.2. A ring in which ACC holds for right as well as left ideals is called a *Noetherian ring*.

EXAMPLE 1. $\mathbb{Z}$ is a Noetherian ring since it is a PID and every PID is a Noetherian ring (Lemma 10.27).

EXAMPLE 2. Every finite ring is clearly both right and left Noetherian.

EXAMPLE 3. Consider a division ring D . Since the only right ideals of D are (0) and D itself, D is a right Noetherian. For similar reasons D is also left Noetherian.

At the end of the chapter we shall give an example of a right Noetherian ring which is not left Noetherian.

Definition 15.3. If in a ring R every non-empty set of right (left) ideals, partially ordered by inclusion, has a maximum element, we say that *maximum condition* holds for right (left) ideals of R .

Notice that by a maximal element of a non-empty set F partially ordered by relation $\leq$, we mean an element $A \in F$ such that there exists no element B in F , satisfying $A < B$. In other words A is a maximal element of F if and only if for each B in F , $A \leq B \Rightarrow A = B$. A set may possess more than one maximal element.

REMARK. As the results for left Noetherian rings are in complete analogy with those for right Noetherian rings, we shall confine ourselves to right Noetherian rings only.

Definition 15.4. A right ideal I of a right R is said to be *finitely generated* if it is generated by a finite subset of R .

NOTATION. If the generating set S consists of n elements $a_1, a_2, \dots, a_n$ we write $I = \langle a_1, a_2, \dots, a_n \rangle_r$. It can be checked that $I = \langle a_1 \rangle_r + \langle a_2 \rangle_r + \dots + \langle a_n \rangle_r$.

Theorem 15.5. For any ring R the following statements are equivalent:—

- (1) R is right Noetherian.
- (2) Maximum condition holds for right ideals of R .
- (3) Every right ideal of R is finitely generated.

Proof: (1) $\Rightarrow$ (2). Let F be a non-empty set of right ideals of R and let I_1 be a right ideal in F . If I_1 is not a maximal element of F we can find a right ideal I_2 of R such that $I_2 \in F$ and $I_1 < I_2$. If F has no maximal element the above process can be continued indefinitely, giving rise to infinite strictly ascending chain

$$I_1 < I_2 < I_3 < \dots$$

of right ideals of R . This contradicts the fact that R is right Noetherian. Hence F has a maximal element, consequently (1) $\Rightarrow$ (2).

(2) $\Rightarrow$ (3). Let I be a right ideal of R and let $F = \{J \mid J \text{ is a finitely generated right ideal of } R \text{ such that } J \subseteq I\}$. Clearly $(0) \in F$ so F is a non-empty set of right ideals. By (2), F has a maximal element. Suppose M is a maximal element of F . Clearly M is finitely generated. We shall show that $I = M$. As $M \in F$, $M \subseteq I$. If possible let $I \neq M$, then there exists an element $a \in I$, $a \notin M$. Consider the right ideal $M + \langle a \rangle_r$. $M + \langle a \rangle_r$ is finitely generated right ideal of R . [It is generated by the set $S \cup \{a\}$ where $\langle S \rangle = M$]. Since $a \in I$, $M + \langle a \rangle_r \subseteq I$. This gives that $M + \langle a \rangle_r \in F$. But $a \notin M \Rightarrow M < M + \langle a \rangle_r$. This violates the maximality of M . Hence $I = M$. As a consequence I is finitely generated.

(3) $\Rightarrow$ (1). Let $I_1 \subseteq I_2 \subseteq I_3 \subseteq \dots$ be an infinite ascending chain of right ideals of R . Consider $I = \bigcup_i I_i$. I is a right ideal of R (Compare Lemma 10.26). Now each right ideal is finitely generated. Let $S = \{a_1, a_2, a_3, \dots, a_n\}$ be a generating set of I , i.e. $I = \langle a_1, a_2, \dots, a_n \rangle_r$.

For each $j \in \{1, 2, 3, \dots, n\}$, $a_j \in I$ implies the existence of some positive integer i_j such that $a_j \in I_{i_j}$. Let $k = \max(i_1, i_2, \dots, i_n)$. Then $a_j \in I_k$ $\forall j = 1, 2, 3, \dots, n$. This gives that $I \subseteq I_k$, but $I_k \subseteq I$, hence $I = I_k$. Now for all $m \geq k$, $I_m \subseteq I = I_k$ gives $I_m = I_k$. Hence R is right Noetherian. ■

Definition 15.6. An ideal I of a ring R is called *right Noetherian* if every strictly ascending chain of right ideals of R , which are contained in I , is finite. By above definition it is clear that every ideal of a right Noetherian ring is right Noetherian.

Lemma 15.7. A homomorphic image of a right Noetherian ring is right Noetherian.

Proof: Let S be a homomorphic image of a right Noetherian ring R . Then $S \cong R/I$ for some ideal I of R . So it is sufficient to prove that R/I is right Noetherian.

Let $J_1 \subseteq J_2 \subseteq J_3 \subseteq \dots$ be an ascending chain of right ideals of R/I . Now each J_i is of the form K_i/I where K_i is a right ideal of R containing I . Also $J_i \subseteq J_{i+1} \Rightarrow K_i \subseteq K_{i+1}$. So the above ascending chain gives rise to ascending chain.

$K_1 \subseteq K_2 \subseteq K_3 \subseteq \dots$, of right ideals of R . But as R is right Noetherian there exists a positive integer n such that $K_m = K_n \forall m \geq n$. This implies that $J_m = J_n \forall m \geq n$. Hence R/I is right Noetherian. ■

Lemma 15.8. (Modular law). If L, M, N are right ideals of a ring R such that $M \subseteq L$, then $L \cap (M + N) = M + (L \cap N)$.

Proof: As M and $L \cap N$ both are contained in L , $M + (L \cap N) \subseteq L$. Further $L \cap N \subseteq N \Rightarrow M + L \cap N \subseteq M + N$. This gives that $M + (L \cap N) \subseteq L \cap (M + N)$. Again let $x \in L \cap (M + N)$, then $x \in L$ and $x = y + z$ for some $y \in M$ and $z \in N$. Now $z = x - y \in L$ as $M \subseteq L$ and this gives that $z \in L \cap N$. Hence $x \in M + (L \cap N)$. So $L \cap (M + N) \subseteq M + (L \cap N)$. Consequently $L \cap (M + N) = M + (L \cap N)$. ■

Theorem 15.9. Let I be an ideal of a ring R then R is right Noetherian if and only if both I and R/I are right Noetherian.

Proof: If R is right Noetherian then by definition 15.6, I is right Noetherian and by Lemma 15.7, R/I is also right Noetherian.

Conversely let both I and R/I be right Noetherian and let $J_1 \subseteq J_2 \subseteq J_3 \subseteq \dots$ be an ascending chain of right ideals of R . Then $J_1 \cap I \subseteq J_2 \cap I \subseteq J_3 \cap I \subseteq \dots$ is an ascending chain of right ideals contained in I . Since I is a right Noetherian there exists a positive integer n such that $J_m \cap I = J_n \cap I \forall m \geq n$.

Again $(J_1 + I)/I \subseteq (J_2 + I)/I \subseteq (J_3 + I)/I \subseteq \dots$ is an ascending chain of right ideals of R/I . As R/I is right Noetherian there exists a positive integer t such that $(J_m + I)/I = (J_t + I)/I \forall m \geq t$. If $r = \max(n, t)$ then $J_m \cap I = J_r \cap I$ and $(J_m + I)/I = (J_r + I)/I \forall m \geq r$. But then $J_m \cap I = J_r \cap I$ and $(J_m + I)/I = (J_r + I)/I \forall m \geq r$. Now for all $m \geq r$, $J_m + I = J_r + I \forall m \geq r$. We claim that $J_m = J_r \forall m \geq r$. Now for all $m \geq r$,

$$\begin{aligned}
 J_m &= J_m \cap (J_m + I) = J_m \cap (J_r + I), \text{ as } J_m + I = J_r + I \forall m \geq r, \\
 &= J_r + J_m \cap I \text{ as } J_r \subseteq J_m \forall m \geq r \text{ and by Modular law,} \\
 &= J_r + J_r \cap I \text{ since } J_m \cap I = J_r \cap I, \\
 &= J_r \text{ as } J_r \cap I \subseteq J_r.
 \end{aligned}$$

Hence $J_m = J_r \forall m \geq r$. As a consequence R is right Noetherian. ■

We now give a proof of a famous theorem of Hilbert.

Theorem 15.10. (Hilbert's Basis Theorem)

If R is a right Noetherian ring with unity then $R[x]$, the ring of polynomials over R , is right Noetherian.

Proof: It is sufficient to prove that every non-zero right ideal of $R[x]$ is finitely generated.

Let I be a non-zero right ideal of $R[x]$. For each integer $k \geq 0$, define $I_k = \{a \in R \mid a \neq 0 \text{ and there exists a polynomial } \alpha_0 + \alpha_1 x + \dots + \alpha_{k-1} x^{k-1} + a x^k \in I\} \cup \{0\}$.

I_k is a right ideal of R and $I_k \subseteq I_{k+1}$ for all integers $k \geq 0$.

Since R is right Noetherian there exists a positive integer n such that $I_m = I_n \forall m \geq n$. Also each I_i being a right ideal of right Noetherian ring R , is finitely generated. Let $I_i = \langle a_{i1}, a_{i2}, a_{i3}, \dots, a_{im_i} \rangle$ for all $i = 0, 1, 2, \dots, n$ where a_{ij} is the leading coefficient of a polynomial $f_{ij} \in I$, of degree i . We claim that I is generated by $m_0 + m_1 + m_2 + \dots + m_n$ polynomials $f_{01}, f_{02}, \dots, f_{0m_0}, f_{11}, f_{12}, \dots, f_{1m_1}, \dots, f_{n1}, f_{n2}, \dots, f_{nm_n}$.

Let $J = \langle f_{01}, f_{02}, \dots, f_{n1}, f_{n2}, \dots, f_{nm_n} \rangle$. As each $f_{ij} \in I$, $J \subseteq I$.

Let $f \neq 0 \in R[x]$ be such that $f \in I$, then $f = c_0 + c_1 x + c_2 x^2 + \dots + c_s x^s$, $c_s \neq 0$. We shall apply induction on s . For $s = 0$, $f = c_0 \in I_0$ and by definition of f_{ij} 's, $a_{01} = f_{01}, a_{02} = f_{02}, \dots, a_{0m_0} = f_{0m_0}$ are elements of I_0 which generate I_0 , so $I_0 \subseteq J$, we get $c_0 \in J$. This gives $f \in J$.

Suppose now that all non-zero polynomials in I of degree $< s$ belong to J and let $\deg f = s$.

Let $s > n$. The leading coefficient c_s of f belongs to I_s .

But $I_s = I_n$ as $I_m = I_n \forall m \geq n$. This implies that $c_s \in I_n$ i.e. $c_s = a_{n1} b_1 + a_{n2} b_2 + a_{n3} b_3 + \dots + a_{nm_n} b_{m_n}$ for some $b_1, b_2, b_3, \dots, b_{m_n} \in R$ (1)

The polynomial $g = f - (f_{n1} b_1 + f_{n2} b_2 + \dots + f_{nm_n} b_{m_n}) x^{s-n}$ is zero or of degree less than s as the coefficient of x^s in g is equal to $c_s - (a_{n1} b_1 + a_{n2} b_2 + \dots + a_{nm_n} b_{m_n})$, which is zero by (1).

If $g = 0$ then $g \in J$, if $g \neq 0$ then by induction hypothesis $g \in J$ and so in each case $f = g + (f_{n1} b_1 + f_{n2} b_2 + \dots + f_{nm_n} b_{m_n}) x^{s-n} \in J$.

If $s \leq n$ then $c_s \in I_s \Rightarrow c_s = a_{s1} d_1 + a_{s2} d_2 + \dots + a_{sm_s} d_{m_s}$ for some $d_1, d_2, \dots$

$\dots, d_{m_s} \in R$.

The polynomial $h = f - (f_{s_1}d_1 + f_{s_2}d_2 + \dots + f_{s_{m_s}}d_{m_s})$ is either zero or of degree $< s$ as the coefficient of x^s in h is equal to $c_s - (a_{s_1}d_1 + a_{s_2}d_2 + \dots + a_{s_{m_s}}d_{m_s}) = 0$. Hence once again $h \in J$, so $f \in J$.

Thus in each case every non-zero polynomial f which is in I is also in J . This gives that $I \subseteq J$, which in turn implies that $I = J$. Consequently I is finitely generated. Hence $R[x]$ is right Noetherian. ■

2. Artinian Rings

Definition 15.11. A ring in which every strictly descending chain of right (left) ideals is finite is called a *Right (Left) Artinian ring*.

Alternatively one can define a *Right (Left) Artinian ring* to be a ring in which for every infinite descending chain of right (left) ideals

$$I_1 \supseteq I_2 \supseteq I_3 \supseteq \dots$$

there exists a positive integer n such that $I_m = I_n \forall m \geq n$.

REMARK. A right (left) Artinian ring is also known as a ring with *DCC* (Descending Chain Condition) on right (left) ideals.

Definition 15.12. A ring in which *DCC* holds for right as well as left ideals is called an *Artinian ring*.

EXAMPLE 4. Every division ring D is right Artinian as its only right ideals are (0) and D itself. Because of similar reason D is also left Artinian.

EXAMPLE 5. Every finite ring is both right and left Artinian.

EXAMPLE 6. Consider the set $\mathbb{Z}(p^\infty)$ of all rational numbers of the form m/p^n between 0 and 1, where p is a fixed prime number, m is an arbitrary positive integer and n runs through all non-negative integers. Then $\mathbb{Z}(p^\infty)$ is an Abelian group under addition modulo 1. We make $\mathbb{Z}(p^\infty)$ a ring by defining $ab = 0 \forall a, b \in \mathbb{Z}(p^\infty)$. Note that each subgroup of $\mathbb{Z}(p^\infty)$ is an ideal of $\mathbb{Z}(p^\infty)$. Now let I be any proper ideal of $\mathbb{Z}(p^\infty)$ and let k be the least positive integer such that for some positive integer a , $a/p^k \in I$. Notice that p does not divide a otherwise we shall have $b/p^{k-1} \in I$ for some positive integer b , which is against the choice of k .

Now $0, 1/p^{k-1}, 2/p^{k-1}, \dots, (p^{k-1}-1)/p^{k-1} \in I$. We contend that I consists of precisely these elements. If our contention is not correct, there exists an element $c/p^i \in I$ where c is a positive integer, $i \geq k$ and p does not divide c . Since $(c, p) = 1$, we can find integers r and s such that $cr + ps = 1$.

Now $(rp^{i-k})c/p^i = rc/p^k$ and $s/p^{k-1} = sp/p^k$, each reduced modulo p lies in I . It follows that $rc/p^k + sp/p^k = (rc + sp)/p^k = 1/p^k \in I$, a contradiction to the choice of k .

Hence ideal $I = \{0, 1/p^{k-1}, 2/p^{k-1}, \dots, (p^{k-1}-1)/p^{k-1}\}$. We denote this ideal by I_{k-1} . Thus we conclude that the only proper ideals of $\mathbf{Z}(p^\infty)$ are of the form I_{j-1} for each positive integer j . Since each proper ideal of $\mathbf{Z}(p^\infty)$ has finite number of elements, DCC holds. Hence $\mathbf{Z}(p^\infty)$ is an Artinian ring.

At the end of the chapter we shall give example of a right Artinian ring which is not left Artinian.

Definition 15.13. If in a ring R , every non-empty collection of all right (left) ideals of R , partially ordered by inclusion relation, has a minimal element we say that *minimum condition* holds for right (left) ideals of R .

Note that by a minimal element of a partially ordered non-empty set F under the relation $\leq$, we mean an element $A \in F$ such that there exists no $B \in F$ satisfying $B < A$. In other words A is a minimal element of F if and only $B \leq A \Rightarrow B = A$. As in the case of maximal element, a set may possess more than one minimal element.

As the results for left Artinian rings are completely analogous to those of right Artinian rings we shall hence forth discuss only right Artinian rings.

Theorem 15.14. For a ring R , the following statements are equivalent.

(1) R is right Artinian.

(2) Minimum condition holds for right ideals of R .

Proof: (1) $\Rightarrow$ (2). Let F be a non-empty set of right ideals of R . Let I_1 be an element of F . If I_1 is not a minimal element, we can find another right ideal I_2 of R such that $I_1 > I_2$. If F has no minimal element this process can be repeated indefinitely giving rise to infinite strictly descending chain $I_1 > I_2 > I_3 > \dots$ of right ideals of R . This is a contradiction to the fact that R is right Artinian. Hence F has a minimal element.

(2) $\Rightarrow$ (1). Let $I_1 \supseteq I_2 \supseteq I_3 \supseteq \dots$ be a descending chain of right ideals of R . Consider $F = \{I_i | i = 1, 2, 3, \dots\}$. Then F is non-empty as $I_1 \in F$. By (2), F has a minimal element, say I_n , for some positive integer n . Now $\forall m \geq n, I_m \subseteq I_n$. If $I_m \neq I_n$ then as I_n is a minimal element of $F, I_m \notin F$, which is absurd. Hence $I_m = I_n \forall m \geq n$. Consequently R is right Artinian. ■

Definition 15.15. An ideal I of a ring R is said to be *right Artinian* if every strictly descending chain of right ideals of R , which are contained in I , is finite.

Clearly every ideal I of a right Artinian ring is right Artinian.

We leave the proofs of the following two theorems to the readers since they are exactly similar to those of the corresponding results

for right Noetherian rings.

Theorem 15.16. *A homomorphic image of right Artinian ring is right Artinian.■*

Theorem 15.17. *Let I be an ideal of a ring R then R is right Artinian if and only if both I and R/I are right Artinian.■*

Theorem 15.18. *A right Artinian ring having more than one element and having no proper zero-divisors is a division ring.*

Proof: R be a right Artinian ring having at least two elements and having no proper zero-divisors. Let $a(\neq 0) \in R$. Consider the descending chain $\langle a \rangle_r \supseteq \langle a^2 \rangle_r \supseteq \langle a^3 \rangle_r \supseteq \dots$, of right ideals of R . As R is right Artinian there exists a positive integer n such that $\langle a^n \rangle_r = \langle a^{n+1} \rangle_r = \langle a^n \rangle_r \Rightarrow a^n \in \langle a^{n+1} \rangle_r = \langle a^n \rangle_r \quad \forall m \geq n$. In particular $\langle a^{n+1} \rangle_r = \langle a^n \rangle_r \Rightarrow a^n \in \langle a^{n+1} \rangle_r$

$$\Rightarrow a^n = a^{n+1}r + ma^{n+1} \text{ for some } r \in R, m \in \mathbb{Z}.$$

$$\Rightarrow a^n = a^n(ar + ma) = a^n e \text{ where } e = ar + ma \in R$$

$$\Rightarrow a = ae \text{ as } a^{n-1}(\neq 0) \text{ is cancelled.}$$

$$\Rightarrow ae = ae^2 \Rightarrow a(e - e)^2 = 0 \Rightarrow e = e^2 \text{ as } a \neq 0 \text{ and } R \text{ is without proper zero-divisors.}$$

Also $e \neq 0$ as $a \neq 0$.

Now for any $x \in R$, $(xe - x)e = 0 \Rightarrow xe = x$ and $e(ex - x) = 0 \Rightarrow ex = x$. Hence e is the unity of R . Consequently $\langle a^n \rangle_r = a^n F = \langle a^{n+1} \rangle_r = a^{n+1} R$. Now $a^n e \in a^n R \Rightarrow a^n e \in a^{n+1} R \Rightarrow a^n e = a^{n+1} b$ for some $b \in R \Rightarrow ab = e$. Hence R is a division ring.■

Corollary 15.19. *An Artinian integral domain with at least two elements is a field.■*

Corollary 15.20. *In an Artinian commutative ring R with unity, every prime ideal different from R is maximal.*

Proof: If P is a prime ideal of R such that $P \neq R$ then R/P is an Artinian integral domain with unity $1+P$. Hence R/P is a field (Corollary 15.19). This gives that P is a maximal ideal of R .■

Theorem 15.21. *In a right Artinian ring every nil right ideal is nilpotent.*

Proof: Let I be a nil right ideal of a right Artinian ring R . For the descending chain of a right ideals $I \supseteq I^2 \supseteq I^3 \supseteq \dots$. there exists a positive integer n such that $I^n = I^{n+m} \quad \forall m \geq n$. In particular $I^{2n} = I^n$. We claim that $I^n = (0)$, if not then $I^{2n} \neq (0)$. Let $F = \{A \mid A \text{ is a right ideal of } R \text{ such that } AI^n \neq (0)\}$. F is non-empty as $I^n \in F$. Since R is right Artinian, F has a minimal element. Let K be a minimal element of F . Then $KI^n \neq (0) \Rightarrow$ there exists $k \in K$ such that $kI^n \neq (0)$. But $kI^n \subseteq K$ as K is a right ideal of R . Let $kI^n < K$. Again $(kI^n)I^n = kI^{2n} = kI^n \neq (0) \Rightarrow kI^n \in F$ as kI^n is a right ideal of R . This violates the minimality of K . Hence $kI^n = K$. Since

$k \in K$, there exists $a \in I^n$ such that $ka = k$. But as I is a nil right ideal and $I^n \subseteq I$, a is nilpotent. So there exists a positive integer t such that $a^t = 0$.

Then $k = ka = ka^2 = ka^3 = \dots = ka^t = 0 \Rightarrow kI^n = (0)$, which is a contradiction to the choice of k .

Hence $I^n = (0)$. Consequently I is a nilpotent right ideal. $\square$

3. Examples and Counter-examples

We close this chapter by giving some interesting examples and counter-examples.

EXAMPLE 7. $\mathbf{Z}$ is Noetherian ring but not Artinian. In fact for any positive integer n , the strictly descending chain.

$$\langle n \rangle \supset \langle 2n \rangle \supset \langle 4n \rangle \supset \dots$$

of ideals of $\mathbf{Z}$ is infinite.

EXAMPLE 8. We have seen in Example 6 that the ring $\mathbf{Z}(p^\infty)$ is Artinian for any fixed prime p . We claim that $\mathbf{Z}(p^\infty)$ is not Noetherian. Recall that every ideal of $\mathbf{Z}(p^\infty)$ is of the type I_{k-1} for every positive integer k . Note also that $x \in I_{k-1} \Rightarrow x = a/p^{k-1}$ for some integer a , $0 \leq a \leq (p^{k-1} - 1) \Rightarrow x = ap/p^k$ and $0 \leq ap \leq p^k - p < p^k - 1 \Rightarrow x \in I_k$. Further by definition of I_{k-1} , $a/p^k \notin I_{k-1}$. So $I_{k-1} \subset I_k$. The strictly ascending chain $(0) \subset I_1 \subset I_2 \subset I_3 \subset \dots$ is infinite. Hence $\mathbf{Z}(p^\infty)$ is not Noetherian. Also notice that although each proper ideal of $\mathbf{Z}(p^\infty)$ is finitely generated, yet $\mathbf{Z}(p^\infty)$ itself is not finitely generated.

EXAMPLE 9. Let S be an infinite set and let R be the set of all subsets of S (i.e. R is the power set of S). Then $(R, \Delta, \cap)$ is a commutative ring with unity.

Recall that for $A, B \in R$, $A \Delta B = (A - B) \cup (B - A)$.

We claim that R is not Noetherian. For this purpose it is sufficient to show that there exists an ideal of R which is not finitely generated. Let C be the subset of R consisting of all finite subsets of S . Clearly C is an ideal of R .

Suppose that C is finitely generated, then $C = \langle C_1, C_2, C_3, \dots, C_t \rangle$ where for each $i = 1, 2, \dots, t$; C_i is a finite subset of S . Let $A \in C$ then $A = (C_1 \cap D_1) \Delta (C_2 \cap D_2) \Delta \dots \Delta (C_t \cap D_t)$ for some subsets $D_1, D_2, \dots, D_t$ of S . Then for each $y \in A$, $y \in (C_1 \cap D_1) \Delta \dots \Delta (C_t \cap D_t)$. This implies that $y \in C_i \cap D_i$ at least for some i , $1 \leq i \leq t$. So each element of A is in some C_i . Now as S is an infinite set and $\bigcup_{i=1}^t C_i$ is a finite subset of S ,

there exists an $x \in S$, $x \notin \bigcup_{i=1}^t C_i$. In other words $x \in S$ and $x \notin C_i$

$\forall i=1, 2, \dots, t$. Consider now the finite subset $\{x\}$ of S . By our choice of C , $\{x\} \in C$. This is absurd in view of the argument given earlier. Hence C is not finitely generated. Consequently R is not Noetherian.

EXAMPLE 10. Let F be the ring of all real valued functions on $\mathbf{R}$. For any real number $r > 0$, we define $I_r = \{f \in F \mid f(x) = 0 \forall -r \leq x \leq r\}$. Then I_r is an ideal of F .

The strictly descending chain of ideals $I_1 > I_{1/2} > I_{1/3} > \dots$ and the strictly ascending chain of ideals $I_1 < I_{1/2} < I_{1/3} < \dots$ never terminate. Hence F is neither Noetherian nor Artinian.

Now we show by examples that a subring of Noetherian (Artinian) ring need not be Noetherian (Artinian).

EXAMPLE 11. Consider $\mathbf{Q}[x]$ as $\mathbf{Q}$ is Noetherian, $\mathbf{Q}[x]$ is Noetherian (Hilbert's Basis Theorem).

Let $R = \{f \in \mathbf{Q}[x] \mid \text{constant term of } f \in \mathbf{Z}\}$.

R is a subring of $\mathbf{Q}[x]$. The strictly ascending chain $\langle x \rangle < \langle x/2 \rangle < \langle x/4 \rangle < \dots$ of ideals of R never terminate.

Hence R is not Noetherian.

EXAMPLE 12. $\mathbf{Q}$, being a field, is Artinian. $\mathbf{Z}$ is a subring of $\mathbf{Q}$, but $\mathbf{Z}$ is not Artinian (see Example 7).

The following example shows that Hilbert's Basis Theorem fails to hold for Artinian rings.

EXAMPLE 13. Let F be a field. Then F is Artinian. Consider $F[x]$. The strictly descending chain

$$\langle x \rangle > \langle x^2 \rangle > \langle x^3 \rangle > \dots$$

of ideals of $F[x]$ is infinite. Hence $F[x]$ is not Artinian.

Observe that as $F[x]$ is an integral domain, if at all $F[x]$ is Artinian we shall get that $F[x]$ is a field by Corollary 15.19.

Finally we give examples of rings which are right Noetherian (Artinian) but not left Noetherian (Artinian).

EXAMPLE 14. Let $R = \left\{ \begin{pmatrix} a & b \\ 0 & c \end{pmatrix} \mid a \in \mathbf{Z}; b, c \in \mathbf{Q} \right\}$.

Then R is a ring under matrix addition and multiplication. We claim that R is right Noetherian but not left Noetherian.

For any non-negative integer k , consider the set

$$A_k = \left\{ \begin{pmatrix} 0 & m/2^k \\ 0 & 0 \end{pmatrix} \mid m \in \mathbf{Z} \right\}.$$

Then A_k is a left ideal of R . Further $A_k < A_{k+1}$ as $m/2^k = 2m/2^{k+1}$

and $\begin{pmatrix} 0 & 1/2^{k+1} \\ 0 & 0 \end{pmatrix} \notin A_k$.

Thus we get a non-terminating strictly ascending chain $A_0 < A_1$

$<A_2 < A_3 < \dots$ of left ideals of R . This shows that R is not left Noetherian.

To prove that R is right Noetherian it is sufficient to establish that each non-zero right ideal of R is finitely generated.

Let $A[\neq(0)]$ be a right ideal of R and let $\alpha e_{11} + \beta e_{12} + \gamma e_{22} \in A$ where e_{ij} denotes the matrix with 1 in $(i, j)^{th}$ position and zero elsewhere, and $\alpha \in \mathbb{Z}$, $\beta, \gamma \in \mathbb{Q}$. Note also that $e_{ij}e_{kl} = e_{il}$ if $j=k$ and $e_{ij}e_{kl} = 0$ if $j \neq k$. Two cases arise.

Case I. $\alpha \neq 0$. Let δ be the least positive integer such that $\delta e_{11} + \beta e_{12} + \gamma e_{22} \in A$ for some $b, c \in \mathbb{Q}$. Then we claim that either A is generated by matrices $\delta e_{11}, e_{12}, e_{22}$ or by δe_{11} and e_{12} . Now $(\delta e_{11} + \beta e_{12} + \gamma e_{22}) e_{12} \in A \Rightarrow \delta e_{12} \in A \Rightarrow \delta e_{12} (1/\delta) e_{22} = e_{12} \in A \Rightarrow \beta e_{12} = e_{12}(\gamma e_{22}) \in A$.

Also $\delta e_{11} + \beta e_{12} + \gamma e_{22} \in A \Rightarrow (\delta e_{11} + \beta e_{12} + \gamma e_{22}) e_{11} \in A \Rightarrow \delta e_{11} \in A$. Thus we get $\gamma e_{22} \in A$. Now in case $\gamma = 0$ for all such δ and b, A is generated by δe_{11} and e_{12} . In the other case, $\gamma e_{22} (1/\gamma) e_{22} \in A$ and A is then generated by matrices $\delta e_{11}, e_{12}$ and e_{22} .

Case II. $\alpha = 0$. In this case $\beta e_{12} + \gamma e_{22} \in A$. If all the elements of A are of the type $\lambda(\beta e_{12} + \gamma e_{22})$ for some $\lambda \in \mathbb{Q}$, then A is generated by single matrix $\beta e_{12} + \gamma e_{22}$. Otherwise there exist $b_1, c_1 \in \mathbb{Q}$ such that $b_1 e_{12} + c_1 e_{22} \in A$ but $b_1 c \neq bc_1$. Then $(bc_1 - b_1 c) e_{12} \in A \Rightarrow (bc_1 - b_1 c) e_{12} [1/(bc_1 - b_1 c)] e_{22} \in A \Rightarrow e_{12} \in A \Rightarrow e_{12} \beta e_{22} \in A \Rightarrow \beta e_{12} \in A \Rightarrow \gamma e_{22} \in A$. Either $c = 0$ or $e_{22} \in A$. Thus A is either generated by e_{12} and e_{22} or by e_{12} alone. Hence A is finitely generated.

Consequently R is right Noetherian.

EXAMPLE 15. Let $R = \left\{ \begin{pmatrix} a & b \\ 0 & c \end{pmatrix} \mid a \in \mathbb{Q}; b, c \in \mathbb{R} \right\}$.

Clearly R is a ring under matrix addition and multiplication. We claim that R is right Artinian but not left Artinian.

R , as a vector space over $\mathbb{Q}$, is of infinite dimension, there exist real numbers $\alpha_1, \alpha_2, \alpha_3, \dots, \alpha_n, \dots$ which are linearly independent over $\mathbb{Q}$. For each positive integer k , put $A_k = \left\{ \begin{pmatrix} 0 & a \\ 0 & 0 \end{pmatrix} \mid a \in \text{subspace of } \mathbb{R} \text{ generated by } \alpha_k, \alpha_{k+1}, \alpha_{k+2}, \dots \right\}$. It can be easily checked that A_k

is a left ideal of R . Further $A_k > A_{k+1}$, as $\alpha_k \notin A_{k+1}$. Since the set $\{\alpha_1, \alpha_2, \alpha_3, \dots\}$ is infinite, we get an infinite strictly descending chain of left ideals $I_1 > I_2 > I_3 > \dots$ of R . This establishes the fact that R is not left Artinian.

Let $A_1 > A_2$ be two right ideals of R and $\alpha e_{11} + \beta e_{12} + \gamma e_{22} \in A_2$ where $\alpha \in \mathbb{Q}$, $\beta, \gamma \in \mathbb{R}$ and e_{ij} 's as defined in previous example. (e_{ij} 's are called *matrix units*). If $\alpha \neq 0$ then as proved in previous example either $\alpha e_{11} \in A_2$, $e_{12} \in A_2$, $e_{22} \in A_2$ or $\alpha e_{11} \in A_2$, $e_{12} \in A_2$ and $\gamma = 0$. In the

former situation $(\alpha e_{11})(1/\alpha)e_{11} \in A_2$ i.e. $e_{11} \in A_2$ and so $A_2 = R$, which is absurd as $A_1 > A_2$. In the latter situation A_2 is generated by e_{11} and e_{12} . Consider now $A_1 > A_2 > A_3$ where A_3 is a right ideal of R . Now let $ae_{11} + be_{12} \in A_3$. If $a \neq 0$ then $(ae_{11} + be_{12})(1/a)e_{11} \in A_3 \Rightarrow e_{11} \in A_3$; so $be_{12} \in A_3$. Now if $b \neq 0$ then this implies $be_{12}(1/b)e_{22} = e_{12} \in A_3$ and thus $A_3 = A_2$, a contradiction. Hence $b = 0$ and consequently A_3 is generated by e_{11} . This at once proves that A_3 is a minimal right ideal of R . Hence the strictly descending chain of right ideals $A_1 > A_2 > A_3 > (0)$ is finite.

Finally if $\alpha = 0$ then $\beta e_{12} + \gamma e_{22} \in A_2$. If all other elements of A_2 are of the type $\lambda(\beta e_{12} + \gamma e_{22})$ for some $\lambda \in R$ then A_2 is minimal right ideal and $A_1 > A_2 > (0)$ is the strictly descending chain of right ideals. Otherwise as seen in previous example A_2 is generated by either e_{12} and e_{22} or by e_{12} only.

In case A_2 is generated by e_{12} only; it is minimal right ideal of R and we are done. In the case A_2 is generated by e_{12} and e_{22} , a similar argument will show that if $A_1 > A_2 > A_3$ for some right ideal A_3 of R , then either A_3 is generated by e_{12} or by $de_{12} + fe_{22}$ for some $d, f \in R$.

In each case A_3 is a minimal right ideal of R . Consequently R does not admit any infinite strictly descending chain of right ideals. This establishes the fact that R is right Artinian.

PROBLEMS

1. Show that if a ring R is Noetherian then every homomorphism of R onto itself is 1-1.
2. Let I be a non-zero ideal of a Principal Ideal Domain R . Prove that R/I is both Artinian and Noetherian.
3. Prove that in a UFD, ACC holds on principal ideals.
4. Prove that every Noetherian ring R with unity has a maximal ideal.

[Hint. Let $F = \{J \mid J \text{ is an ideal of } R \text{ and } 1 \notin J\}$.]

5. If R is commutative ring with unity prove that if $R[x]$ is Noetherian (Artinian) then R is Noetherian (Artinian).

[Hint. $R[x]/(x) \cong R$.]

6. Prove that if an integral domain with unity has finite number of ideals, then it is a field.
7. Prove that every commutative Artinian ring possesses a finite number of proper prime ideals.
8. Prove that every ideal of a commutative Noetherian ring contains a product of prime ideals.
9. Let $R_1, R_2, \dots, R_n$ be n rings and let R be the set of n -tuple

$(a_1, a_2, \dots, a_n)$ with $a_i \in R_i \forall i=1, 2, \dots, n$. Show that R is a ring under component-wise addition and multiplication. R is called the *External Direct Sum* of $R_1, R_2, \dots, R_n$ and we write $R=R_1 \oplus R_2 \oplus \dots \oplus R_n$.

10. Let R be the direct sum of rings $R_1, R_2, \dots, R_n$. For each $i=1, 2, \dots, n$ put $R_i' = \{(0, 0, \dots, 0, \underbrace{a_i}_{\text{ith place}}, \dots, 0) \mid a_i \in R_i\}$. Prove that (i) Each R_i' is an ideal of R . (ii) For each $i, R_i' \cong R_i$.

11. If R is a direct sum of two rings R_1 and R_2 prove that $R/R_1' \cong R_2$ and $R/R_2' \cong R_1$ where R_1' and R_2' are defined as in previous problem.

12. Let R be the direct sum of rings $R_1, R_2, \dots, R_n$. Show that R is right Noetherian (Artinian) if and only if each R_i is right Noetherian (Artinian).

[Hint. Apply induction on n and use Theorem 15.9].

13. If R is right Noetherian (Artinian) prove that R_n , the ring of $n \times n$ matrices over R is right Noetherian (Artinian).

14. (Gilmer) If R is a commutative ring such that $R[x]$ is Noetherian, then show that R contains unity.

15. If in a Noetherian ring R every ideal generated by two elements is principal show that R is a Principal Ideal Ring.

16. Let R be set of all element of type $\frac{f(t)}{g(t)}$ where $f(t), g(t)$ are polynomials in $Z_2(t)$, where Z_2 is the ring of integers modulo 2 and $g(t) \neq 0$. Show that R is Artinian ring under usual addition and multiplication of rational functions and $\theta: R \rightarrow R$ given by $\theta(x) = x^2$ is 1-1 homomorphism which is not onto.

REMARK. This shows that dual of result stated in problem 1 above, is not necessarily true for Artinian rings

CANONICAL FORMS

In Section 4 of Chapter 12 we proved that corresponding to each choice of a basis of a vector space V over a field F , there is a unique matrix associated with a given linear transformation on V . For practical purposes it is often desirable to choose a basis so that the corresponding matrix is in a "nice form". By nice form we mean a matrix which mostly resembles a diagonal matrix. Such forms are readily amenable to further algebraic treatments. In this chapter we discuss the problems of choosing such bases of V for a given $T \in A(V)$.

1. Minimal Polynomials and Characteristic Roots

Definition 16.1. $T \in A(V)$ is called *right (left) invertible* if there exists $S \in A(V)$ such that $TS = I$ ($ST = I$) where I is the identity transformation on V .

Note. S is called a *right (left) inverse* of T .

Definition 16.2. $T \in A(V)$ is called *invertible or regular* if T is both right and left invertible.

Lemma 16.3. If $T \in A(V)$ is regular then there exists $S \in A(V)$ such that $ST = TS = I$.

Proof: Since T is both right and left invertible there exist $S', S'' \in A(V)$ such that $TS' = I$ and $S''T = I$. Now $S' = IS' = (S''T)S' = S''(TS') = S''I = S''$. Putting $S = S' = S''$ we get that $TS = ST = I$. ■

REMARK. It can be shown that if $T \in A(V)$ is regular then there exists *unique* S such that $TS = ST = I$. S is called *inverse* of T and we denote it by T^{-1} .

EXAMPLE 1. Let $V = \mathbb{R}^2$. Consider the basis $\{e_1, e_2\}$ of V where $e_1 = (1, 0)$ and $e_2 = (0, 1)$. Define $T \in A(V)$ by $T(e_1) = 7e_1$ and $T(e_2) = 3e_2$. We claim that T is regular. Let $S \in A(V)$ be defined by $S(e_1) = \frac{e_1}{7}$ and $S(e_2) = \frac{e_2}{3}$. Then it can be easily verified that $ST = TS = I$. Hence T is regular.

EXAMPLE 2. Let $V = \mathbf{R}[x]$, the vector space over polynomials with real coefficients. Define $S, T \in A(V)$ by $S[f(x)] = \frac{d}{dx} [f(x)]$ and $T[f(x)] = \int_1^x f(x) dx$ for all $f(x) \in \mathbf{R}[x]$. Now $ST[f(x)] = S \left[\int_1^x f(x) dx \right] = f(x) \Rightarrow ST = I$. This gives that S is right invertible.

Again $TS[f(x)] = T \left[\frac{d}{dx} f(x) \right] = \int_1^x \left[\frac{df(x)}{dx} \right] dx = \left[f(x) \right]_1^x = f(x) - f(1)$. Since in general $f(1) \neq 0$, $TS \neq I$. Hence S is not regular. [Cf. Lemma 16.3 and its remarks].

Definition 16.4. $T \in A(V)$ is said to be *equivalent* to $S \in A(V)$ if there exist regular transformations $P, Q \in A(V)$ such that $T = PSQ$.

Definition 16.5. $T \in A(V)$ is said to be *similar* to $S \in A(V)$ if there exists a regular transformation $P \in A(V)$ such that $T = P^{-1}SP$.

Throughout the remaining portion of the chapter (unless otherwise stated) V will be a vector space of dimension $n > 0$ over a field F .

It was shown in Theorem 12.17 that $A(V) = \text{Hom}_F(V, V)$ is an algebra of dimension n^2 over F . Let $g(x) = \alpha_0 + \alpha_1 x + \dots + \alpha_k x^k \in F[x]$. For a given $T \in A(V)$, $\alpha_0 I + \alpha_1 T + \dots + \alpha_k T^k \in A(V)$. $g(T)$ will denote $\alpha_0 I + \alpha_1 T + \dots + \alpha_k T^k$.

Definition 16.6. $T \in A(V)$ is called a *root* of a polynomial $g(x) \in F[x]$ if $g(T) = \bar{0}$, the zero linear transformation on V .

Note. For the sake of convenience, we shall use O or 0 in place of $\bar{0}$, but it will clear from reference and context whether O or 0 stands for zero vector or for zero linear transformation, or for zero scalar.

Theorem 16.7. For any non-zero $T \in A(V)$ there exists a unique monic polynomial $m(x) \in F[x]$ such that

- (i) $m(T) = 0$.
- (ii) For any $g(x) \in F[x]$, $g(T) = 0$ if and only if $m(x)$ divides $g(x)$.
- (iii) $F[T] \cong \frac{F[x]}{\langle m(x) \rangle}$.

Proof: Let $F[T] = \{g(T) \mid g(x) \in F[x]\}$. It can be easily seen that $F[T]$ is a subring of $A(V)$ and $\sigma: F[x] \rightarrow F[T]$ defined by $\sigma[g(x)] = g(T)$ is a ring epimorphism. Hence $F[T] \cong \frac{F[x]}{\text{Ker } \sigma}$.

Now $F[x]$ is a PID, so $\text{Ker } \sigma = \langle p(x) \rangle$ for some $p(x) \in F[x]$. As $A(V)$ is of dimension n^2 ; $I, T, T^2, \dots, T^{n^2}$ are L.D. over F . So there exist α_i 's ($0 \leq i \leq n^2$) in F , not all zero such that $\alpha_0 I + \alpha_1 T + \dots + \alpha_{n^2} T^{n^2} = 0$. Hence the non-zero polynomial $f(x) = \alpha_0 + \alpha_1 x + \dots + \alpha_{n^2} x^{n^2} \in \text{Ker } \sigma$. So $\text{Ker } \sigma \neq (0)$ and as a consequence $p(x) \neq 0$; also $p(x)$ is not a constant polynomial as $T \neq 0$. Hence $\deg p(x) > 0$. So we can write

$p(x) = \alpha m(x)$ where α is the leading coefficient of $p(x)$ and $m(x)$ is a monic polynomial. Then $\text{Ker } \sigma = \langle m(x) \rangle$. Now by definition of $\text{Ker } \sigma$, $m(T) = 0$ and for any $g(x) \in F[x]$, $g(x) = 0 \Leftrightarrow g(x) \in \text{Ker } \sigma = \langle m(x) \rangle \Leftrightarrow m(x) \mid g(x)$.

Hence $m(x)$ satisfies (i), (ii) and (iii). Let $m'(x)$ be another monic polynomial over F satisfying (i) and (ii). Then by (ii), $m'(x) \mid m(x)$ and $m(x) \mid m'(x)$. Consequently $m'(x) = \beta m(x)$ for some $\beta (\neq 0) \in F[x]$. Comparing the leading coefficients on both sides and noting that $m(x)$ and $m'(x)$ both are monic we get that $\beta = 1$. So that $m(x) = m'(x)$. This proves the uniqueness of $m(x)$. ■

Definition 16.8. Let $T (\neq 0) \in A(V)$. The monic polynomial $m(x) = x^k - \alpha_1 x^{k-1} - \dots - \alpha_k$ over F is called the *minimal polynomial* of T if (i) $m(T) = T^k - \alpha_1 T^{k-1} - \dots - \alpha_k I = 0$ and (ii) for any $g(x) \in F[x]$, $g(T) = 0$ if and only if $m(x)$ divides $g(x)$.

Theorem 16.7 show that the minimal polynomial of least degree of transformation T on V is the monic polynomial of least degree of which T is a root. In the proof of this theorem it was seen that T is root of a polynomial of degree not exceeding n . Hence the degree of the minimal polynomial of T cannot exceed n . Indeed we shall see that it cannot exceed n .

Theorem 16.9. Let $T (\neq 0) \in A(V)$ and $S \in A(V)$ be invertible. Then T and $S^{-1}TS$ have same minimal polynomials.

Proof: New for any $g(x) = \alpha_0 + \alpha_1 x + \dots + \alpha_k x^k$ in $F[x]$, $g(S^{-1}TS) = \alpha_0 I + \alpha_1 (S^{-1}TS) + \alpha_2 (S^{-1}TS)^2 + \dots + \alpha_k (S^{-1}TS)^k$. However $(S^{-1}TS)^m = S^{-1}T^m S$ for all integers $m \geq 0$. Hence $g(S^{-1}TS) = \alpha_0 I + \alpha_1 S^{-1}TS + \alpha_2 S^{-1}T^2 S + \dots + \alpha_k S^{-1}T^k S = S^{-1}(\alpha_0 I + \alpha_1 T + \alpha_2 T^2 + \dots + \alpha_k T^k)S = S^{-1}g(T)S$. Hence $g(T) = 0 \Leftrightarrow g(S^{-1}TS) = 0$. So that if $m(x)$ and $m'(x)$ are the minimal polynomials of T and $S^{-1}TS$ respectively then $m(T) = 0$, $m'(T) = 0$, $m(S^{-1}TS) = 0$. The definition of minimal polynomial yields $m(x) \mid m'(x)$ and $m'(x) \mid m(x)$. Hence $m(x) = m'(x)$, since both are monic. This proves the result. ■

Let $(v_1, v_2, \dots, v_n)$ be an ordered basis of V . Then $A(V) \cong M_n(F)$, the ring of $n \times n$ matrices over F under the mapping $\sigma : A(V) \rightarrow M_n(F)$ such that $\sigma(T) = (\alpha_{ij}) \in M_n(F)$ where (α_{ij}) is the matrix of T relative to the ordered basis $(v_1, v_2, \dots, v_n)$. Since σ is an isomorphism it follows that for $g(x) \in F[x]$, $g(T) = 0$ if and only if $g(A) = O$ where $A = (\alpha_{ij})$. Hence in particular $m(A) = O$, where $m(x)$ is the minimal polynomial of T .

Definition 16.9 (a). We say that $m(x)$ is also the minimal polynomial of the matrix A .

Theorem 16.10. A linear transformation $T : V \rightarrow V$ is regular if and only if its minimal polynomial has non-zero constant term.

Proof: Let $m(x) = x^k - \alpha_1 x^{k-1} - \alpha_2 x^{k-2} - \dots - \alpha_k$ be the minimal polynomial of T . Then $T^k - \alpha_1 T^{k-1} - \dots - \alpha_{k-1} T - \alpha_k I = 0$ yields $T(T^{k-1} - \alpha_1 T^{k-2} - \dots - \alpha_{k-1} I) = (T^{k-1} - \alpha_1 T^{k-2} - \dots - \alpha_{k-1} I)T = \alpha_k I \dots (1)$.

Let T be regular. If $\alpha_k = 0$, then as T^{k-1} exists, (1) implies that $T^{k-1} - \alpha_1 T^{k-2} - \dots - \alpha_{k-1} I = 0$.

This contradicts the fact that $\deg m(x) = k$ and T cannot be a root of any non-zero polynomial over F of degree $< k$. Hence $\alpha_k \neq 0$.

Conversely if $\alpha_k \neq 0$, then (1) yields that $TT' = T'T = I$ where $T' = \alpha_k^{-1}(T^{k-1} - \alpha_1 T^{k-2} - \dots - \alpha_{k-1} I)$. Hence T is regular. ■

Corollary 15.11. If $T \in A(V)$ is regular then T^{-1} is a polynomial in T over F .

Proof: If $x^k - \alpha_1 x^{k-1} - \alpha_2 x^{k-2} - \dots - \alpha_k$ is the minimal polynomial of T over F then as we have seen in the proof of above theorem, $\alpha_k \neq 0$ and $T^{-1} = \alpha_k^{-1}(T^{k-1} - \alpha_1 T^{k-2} - \dots - \alpha_{k-1} I)$, polynomial in T over F . ■

EXAMPLE 3. Consider $V = \mathbb{Q}^3$. Let (e_1, e_2, e_3) be its standard basis. Let $T \in A(V)$ have the corresponding matrix with respect to

standard basis, given by $A = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix}$. Then $A - 2I = 0$ where

$I = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$. Hence $T - 2I = 0$. This shows that T is a root of

the monic polynomial $x - 2$. Since the minimal polynomial of T is a monic polynomial of least positive degree, of which T is a root, we get that $x - 2$ is the minimal polynomial of T .

EXAMPLE 4. Consider linear transformation $T: \mathbb{Q}^3 \rightarrow \mathbb{Q}^3$ such that $T(e_1) = e_2$, $T(e_2) = e_3$, $T(e_3) = e_1$, where (e_1, e_2, e_3) is the standard basis of $\mathbb{Q}^3$. The matrix of T with respect to this basis is

$$A = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}, A^2 = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}, A^3 = I = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \text{So } T^3 - I = 0.$$

Hence the minimal polynomial $m(x)$ of T divides $x^3 - 1$. So $\deg m(x) \leq 3$. If $\deg m(x) = 1$, then $m(x) = x - \alpha$ for some $\alpha \in \mathbb{Q}$ gives $A - \alpha I = 0$, i.e., $A = \alpha I$, a diagonal matrix. This is a contradiction. Hence $\deg m(x) > 1$. If $\deg m(x) = 2$, then $m(x) = x^2 - \alpha x - \beta$ for some $\alpha, \beta \in \mathbb{Q}$. Hence $A^2 - \alpha A - \beta I = 0$ implies that

$$\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} - \begin{pmatrix} 0 & 0 & \alpha \\ \alpha & 0 & 0 \\ 0 & \alpha & 0 \end{pmatrix} - \begin{pmatrix} \beta & 0 & 0 \\ 0 & \beta & 0 \\ 0 & 0 & \beta \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

i.e.
$$\begin{pmatrix} -\beta & 1 & -\alpha \\ -\alpha & -\beta & 1 \\ 1 & -\alpha & -\beta \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

Hence $1=0$. This is absurd! Hence $\deg m(x) > 2$. So as $m(x) \mid x^3-1$, we get $\deg m(x) = 3$ and $m(x) = x^3-1$. Thus x^3-1 is the minimal polynomial of T over $\mathbb{Q}$.

EXAMPLE 5. Consider the vector space V of all polynomials over $\mathbb{R}$ of degree not exceeding three. Let $D: V \rightarrow V$ be a differentiation operator. D is a linear transformation (Chapter 12, Example 1). Now $(1, x, x^2, x^3)$ is an ordered basis of V . As $D(1)=0$, $D(x)=1$, $D(x^2)=2x$ and $D(x^3)=3x^2$, the matrix of D relative to this ordered

basis is given by $A = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 \end{pmatrix}$

It can be checked that $A^4=0$, $A^3 \neq 0$. So the minimal polynomial of D divides y^4 . Since $A^3 \neq 0$ and only non-constant factors of y^4 are y, y^2, y^3, y^4 , we get that y^4 is the minimal polynomial of D .

Definition 16.12. Let $T: V \rightarrow V$ be a linear transformation:

(i) Any $v \in V$ is called an *eigen vector* of T if $T(v) = \alpha v$ for some $\alpha \in F$.

(ii) Any $\alpha \in F$ is called an *eigen value* of T if $T(v) = \alpha v$ for some non-zero $v \in V$.

(iii) For any eigen value α of T , any $v \in V$ is called an *eigen vector* of T belonging to α if $T(v) = \alpha v$; the set of all eigen vectors of T belonging to an eigen value α is called *Eigen Space* belonging to α .

Let A be an $n \times n$ matrix over F and $T: V \rightarrow V$ be a linear transformation such that A is the matrix of T relative to some ordered basis $B = (v_1, v_2, \dots, v_n)$ of V . Then by an eigen value of A we mean an eigen value of T , and an n -tuple $(\alpha_1, \alpha_2, \dots, \alpha_n)$ in $F^{(n)}$ is called an eigen vector of A if $\alpha_1 v_1 + \alpha_2 v_2 + \dots + \alpha_n v_n$ is an eigen vector of T . Eigen vectors and eigen values are also called *characteristic vectors* and *characteristic values* (or *roots*) respectively.

Lemma 16.13. (a) If v is a non-zero eigen vector of $T \in A(V)$ then $\alpha \in F$ such that $T(v) = \alpha v$ is unique.

(b) For any eigen value α of T , its eigen space is a non-zero subspace of V .

Proof: (a) Let $T(v) = \alpha v$ and $T(v) = \alpha' v$ for some $\alpha, \alpha' \in F$. Then $(\alpha - \alpha')v = 0$. As $v \neq 0$, we get $\alpha - \alpha' = 0$, i.e., $\alpha = \alpha'$ (Lemma 11.3). This proves the uniqueness of α .

(b) Let W_α be the eigen space of α . By definition of eigen value, there exists $v(\neq 0) \in V$ such that $T(v) = \alpha v$. Thus $v \in W_\alpha$ and hence $W_\alpha \neq (0)$. Let $v, w \in W_\alpha$, $\lambda, \mu \in F$. Then $T(\lambda v + \mu w) = \lambda T(v) + \mu T(w) = \lambda(\alpha v) + \mu(\alpha w) = \alpha(\lambda v + \mu w)$. Hence $\lambda v + \mu w \in W_\alpha$. This proves that W_α is a subspace of V (Theorem 11.6). ■

Definition 16.14. For each $\alpha \in F$, dimension of W_α , the eigen space of α , is called the *geometric multiplicity* of α .

EXAMPLE 6. Let $T: Q \rightarrow Q$ be a L.T., such that its matrix with respect to the standard basis (e_1, e_2, e_3) is $A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$. Then

$T(e_1) = e_1 = 1e_1$, $T(e_2) = e_2 = 1e_2$ and $T(e_3) = 0 = 0e_3$. Thus e_1, e_2, e_3 are all eigen vectors and 1, 1, 0 are the corresponding eigen values; although e_1, e_2, e_3 are L.I., 1, 1, 0 are not distinct. Let us determine eigen spaces. Now any $\alpha e_1 + \beta e_2 + \gamma e_3 \in W_1$, the eigen space belonging to 1 if and only if $T(\alpha e_1 + \beta e_2 + \gamma e_3) = 1(\alpha e_1 + \beta e_2 + \gamma e_3)$
 $\Leftrightarrow \alpha T(e_1) + \beta T(e_2) + \gamma T(e_3) = \alpha e_1 + \beta e_2 + \gamma e_3$
 $\Leftrightarrow \alpha e_1 + \beta e_2 = \alpha e_1 + \beta e_2 + \gamma e_3 \Leftrightarrow \gamma = 0$.

Hence W_1 is the subspace spanned by $\{e_1, e_2\}$. Similarly the eigen space belonging to 0 is the subspace spanned by $\{e_3\}$.

Theorem 16.14. Let $T(\neq 0) \in A(V)$.

(a) Let v be an eigen vector of T and $\alpha \in F$ be such that $T(v) = \alpha v$. Then for any $f(x) \in F[x]$, v is an eigen vector of $f(T)$ and $[f(T)](v) = f(\alpha)v$.

(b) Let $\alpha_1, \alpha_2, \dots, \alpha_m$ be distinct eigen values of T and $v_1, v_2, \dots, v_m$ be non-zero eigen vectors of T belonging to $\alpha_1, \alpha_2, \dots, \alpha_m$ respectively, then $v_1, v_2, \dots, v_m$ are L.I. over F .

Proof: (a) Now $T(v) = \alpha v \Rightarrow T^2(v) = T[T(v)] = T(\alpha v) = \alpha T(v) = \alpha(\alpha v) = \alpha^2 v$. $T^3(v) = T[T^2(v)] = T(\alpha^2 v) = \alpha^2 T(v) = \alpha^2(\alpha v) = \alpha^3 v$. In general for any positive integer k , $T^k(v) = \alpha^k v$. Then for any $f(x) = \beta_0 + \beta_1 x + \dots + \beta_m x^m \in F[x]$, $[f(T)](v) = (\beta_0 I + \beta_1 T + \dots + \beta_m T^m)(v) = \beta_0 v + \beta_1 T(v) + \beta_2 T^2(v) + \dots + \beta_m T^m(v) = \beta_0 v + \beta_1(\alpha v) + \beta_2(\alpha^2 v) + \dots + \beta_m(\alpha^m v) = (\beta_0 + \beta_1 \alpha + \dots + \beta_m \alpha^m)(v) = f(\alpha)v$. This proves (a).

(b) If $m=1$, the result is immediate since any non-zero vector is always L.I. over F . We apply induction on m . So let $m>1$ and the result holds for $m-1$ eigen vectors. By induction hypothesis $v_1, v_2, \dots, v_{m-1}$ are L.I. over F and also $v_2, v_3, \dots, v_m$ are L.I. Suppose on the contrary; $v_1, v_2, \dots, v_m$ be L.D. over F . There exist β_i 's ($1 \leq i \leq m$) not all zero such that

$$\beta_1 v_1 + \beta_2 v_2 + \dots + \beta_m v_m = 0$$

If $\beta_m = 0$ then the linear independence of $\{v_i\}_{1 \leq i \leq m-1}$ will yield that

$\beta_i = 0 \quad \forall \quad i = 1, 2, \dots, m-1$. This contradicts the fact that β_i 's ($1 \leq i \leq m$) are not all zero. So suppose that $\beta_m \neq 0$. Operating T on (1) we get $\beta_1 T(v_1) + \beta_2 T(v_2) + \dots + \beta_m T(v_m) = 0$. However $T(v_i) = \alpha_i v_i$ ($1 \leq i \leq m$). Hence

$$\beta_1 \alpha_1 v_1 + \beta_2 \alpha_2 v_2 + \dots + \beta_m \alpha_m v_m = 0 \quad \dots (2)$$

Multiplying (1) by α_1 and then subtracting it from (2), we get that

$$\beta_2 (\alpha_2 - \alpha_1) v_2 + \dots + \beta_m (\alpha_m - \alpha_1) v_m = 0 \quad \dots (3)$$

As $\beta_m \neq 0$ and $\alpha_m \neq \alpha_1$, the coefficient $\beta_m (\alpha_m - \alpha_1)$ of v_m is non-zero. Then by (3); $v_2, v_3, \dots, v_m$ are L.D. This contradicts that $v_2, v_3, \dots, v_m$ are L.I. Hence $v_1, v_2, \dots, v_m$ are L.I. ■

Corollary 16.15. A linear transformation T of V cannot have more than $n (= \dim V)$ distinct eigen values.

Proof: As seen in the above theorem, given any m distinct eigen values of T , corresponding m eigen vectors are L.I. However as $\dim V = n$, V cannot have more than n L.I. vectors. Consequently $m \leq n$. ■

Corollary 16.16. Let $\alpha_1, \alpha_2, \dots, \alpha_m$ be distinct eigen values of a linear transformation $T: V \rightarrow V$. If $W_1, W_2, \dots, W_m$ are subspaces belonging to $\alpha_1, \alpha_2, \dots, \alpha_m$ respectively, then the sum $W_1 + W_2 + \dots + W_m$ is direct.

Proof: Let $v_1 + v_2 + \dots + v_m = 0$ for some $v_i \in W_i$ ($1 \leq i \leq m$). We claim that all v_i 's are zero. Suppose that our claim is not true. Ignoring those v_i 's which are equal to zero, we get

$$v_{i_1} + v_{i_2} + \dots + v_{i_t} = 0 \quad \dots (1)$$

where v_{i_j} ($1 \leq j \leq t$) are all non-zero vectors. Hence these v_{i_j} 's are L.D., although the corresponding eigen values α_{i_j} 's ($1 \leq j \leq t$) are all distinct. This contradicts the above theorem. Hence all the v_i 's are zero. This proves that $W_1 + W_2 + \dots + W_m$ is direct. ■

Theorem 16.17. If a linear transformation $T: V \rightarrow V$ has n distinct eigen values $\alpha_1, \alpha_2, \dots, \alpha_n$ then V has an ordered basis $(v_1, v_2, \dots, v_n)$ such that the matrix of T relative to this basis is

$$\begin{pmatrix} \alpha_1 & 0 & 0 & \dots & 0 \\ 0 & \alpha_2 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & \alpha_n \end{pmatrix}$$

Proof: Let v_i be an eigen vector of T belonging to α_i ($1 \leq i \leq n$). By Theorem 16.14, $v_1, v_2, \dots, v_n$ are L.I. over F . Hence $\mathbf{B} = (v_1, v_2, \dots, v_n)$ is an ordered basis of V (Corollary 11.24).

Since $T(v_i) = \alpha_i v_i \quad \forall i = 1, 2, \dots, n$; we get that the matrix of T relative to $\mathbf{B}$ is equal to

$$\begin{pmatrix} \alpha_1 & 0 & 0 & \dots & 0 \\ 0 & \alpha_2 & 0 & \dots & 0 \\ 0 & 0 & \alpha_3 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & \alpha_n \end{pmatrix}$$

This proves the result. ■

Definition 16.18. $T \in A(V)$ is said to be *singular* (non-singular) if $\text{Ker } T$ is non-zero (zero).

REMARK. For a finite-dimensional vector space T is non-singular if and only if T is regular.

The following theorem is of significance in the determination of eigen values.

Theorem 16.19. $\alpha \in F$ is an eigen value of $T \in A(V)$ if and only if $T - \alpha I$ is singular.

Proof: Let α be an eigen value of T . There exists $v (\neq 0) \in V$ such that $T(v) = \alpha v$. This gives that $(T - \alpha I)v = T(v) - \alpha v = 0$ and so $v (\neq 0) \in \text{Ker } (T - \alpha I)$. Hence $T - \alpha I$ is singular.

Conversely if $T - \alpha I$ is singular, $\text{Ker } (T - \alpha I) \neq (0)$ yields that there exists $v (\neq 0) \in \text{Ker } (T - \alpha I)$. Then $(T - \alpha I)v = 0 \Rightarrow T(v) = \alpha v \Rightarrow \alpha$ is an eigen value of T . ■

Corollary 16.20. If $T \in A(V)$ is singular then 0 is always an eigen value of T .

Proof: Since $T - 0I = T$ is singular, the result follows from the above theorem. ■

We shall show later that every root in F of the minimal polynomial of a linear transformation T is an eigen value of T . Here we prove the converse of the above statement.

Theorem 16.21. Every eigen value of $T \in A(V)$ is a root of the minimal polynomial of T .

Proof: Let $m(x)$ be the minimal polynomial of T . Let $\alpha \in F$ be an eigen value of T . Then for some non-zero $v \in V$, $T(v) = \alpha v$. By Lemma 16.13, $[f(T)](v) = f(\alpha)v$ for every $f(x) \in F[x]$. In particular $[m(T)]v = m(\alpha)v$. This yields that $m(\alpha)v = 0$, since $m(T) = 0$. Hence $m(\alpha) = 0$ as $v \neq 0$. Thus α is a root of the minimal polynomial of T . ■

Now we discuss some additional examples.

EXAMPLE 7. Let V be the vector space of all polynomials over $\mathbb{Q}$ of degree not exceeding 3. Let D be the differentiation operator on V (Example 5). Suppose that $f(x) = \alpha_0 + \alpha_1 x + \alpha_2 x^2 + \alpha_3 x^3$ is an eigen vector of D . Then for some $\alpha \in \mathbb{Q}$, $Df(x) = \alpha f(x)$.
i.e. $\alpha_1 + 2\alpha_2 x + 3\alpha_3 x^2 = \alpha(\alpha_0 + \alpha_1 x + \alpha_2 x^2 + \alpha_3 x^3)$.
Comparing coefficients, we get $\alpha_1 = \alpha\alpha_0$, $2\alpha_2 = \alpha\alpha_1$, $3\alpha_3 = \alpha\alpha_2$.

$0 = \alpha x_3$. So if $\alpha \neq 0$ then these equations will give $\alpha_3 = 0 = \alpha_2 = \alpha_1 = \alpha_0$, i.e., $f(x) = 0$. Hence no non-zero rational number is an eigen value of D . Now if $\alpha = 0$, we get from the above equations, $\alpha_1 = 0$, $\alpha_2 = 0$, $\alpha_3 = 0$. Hence $f(x) = \alpha_0$, a constant. Thus every constant polynomial is an eigen vector of D .

Now $D^2 f(x) = 2\alpha_2 + 6\alpha_3 x$. If $f(x)$ is an eigen vector of D^2 then $D^2 f(x) = \alpha f(x)$ for some $\alpha \in F$; i.e., $2\alpha_2 + 6\alpha_3 x = \alpha(\alpha_0 + \alpha_1 x + \alpha_2 x^2 + \alpha_3 x^3)$. Comparing the coefficients we get $2\alpha_2 = \alpha\alpha_0$, $6\alpha_3 = \alpha\alpha_1$, $0 = \alpha\alpha_2$, $0 = \alpha\alpha_3$. If $\alpha \neq 0$, then these equations yield $f(x) = 0$ (as before). If $\alpha = 0$, we get $\alpha_3 = \alpha_2 = 0$. So that $f(x) = \alpha_0 + \alpha_1 x$. Thus D^2 has no non-zero eigen value and the only non-zero polynomials which are eigen vectors of D are of degree ≤ 1 .

EXAMPLE 8. Consider $V = \mathbb{R}^2$. Let (e_1, e_2) be its standard basis and let $T \in A(V)$ be given by $T(e_1) = e_1 + e_2$ and $T(e_2) = e_1 - e_2$. Suppose that $v = \alpha_1 e_1 + \alpha_2 e_2$ is an eigen vector of T then for some $\alpha \in F$, $T(\alpha_1 e_1 + \alpha_2 e_2) = \alpha(\alpha_1 e_1 + \alpha_2 e_2)$, i.e., $\alpha_1(e_1 + e_2) + \alpha_2(e_1 - e_2) = \alpha\alpha_1 e_1 + \alpha\alpha_2 e_2$. Comparing the coefficients of e_1, e_2 we obtain

$$\alpha_1 + \alpha_2 = \alpha\alpha_1 \quad \text{i.e.} \quad \alpha_1(1 - \alpha) + \alpha_2 = 0 \quad \dots (1)$$

$$\alpha_1 - \alpha_2 = \alpha\alpha_2 \quad \text{i.e.} \quad \alpha_1 - \alpha_2(1 + \alpha) = 0 \quad \dots (2)$$

Multiplying (2) by $(1 - \alpha)$ and subtracting from (1) we get

$$\alpha_2(2 - \alpha^2) = 0$$

Thus if $\alpha \neq \pm \sqrt{2}$ then $\alpha_2 = 0$ and (2) yields $\alpha_1 = 0$. So that $v = 0$. Hence if $\alpha = \sqrt{2}$, (2) yields $\alpha_2 = \frac{\alpha_1}{1 + \sqrt{2}}$ so that in particular taking $\alpha_1 = 1$, we get that $v = e_1 + \frac{1}{1 + \sqrt{2}} e_2$ is a non-zero eigen vector of such that $T(v) = \sqrt{2}v$. Hence $\sqrt{2}$ is an eigen value of T and $v_1 = e_1 + \frac{1}{1 + \sqrt{2}} e_2$ is an eigen vector belonging to $\sqrt{2}$. If $\alpha = -\sqrt{2}$ then $\alpha_2 = \frac{1}{1 - \sqrt{2}} \alpha_1$. In particular we get $-\sqrt{2}$ is also an eigen value of T and $v_2 = e_1 - \frac{1}{1 - \sqrt{2}} e_2$ is an eigen vector of T belonging to $-\sqrt{2}$. Thus T has two distinct eigen values, viz. $\sqrt{2}$, $-\sqrt{2}$. It cannot have any other eigen value because of Corollary 16.15 and the fact that $\dim V = 2$. We leave it to the readers to verify that $\{v_1, v_2\}$ is a L.I. subset and hence $B = (v_1, v_2)$ is an ordered basis of V . As $T(v_1) = \sqrt{2}v_1$, $T(v_2) = -\sqrt{2}v_2$. So the matrix of T relative to B is $\begin{pmatrix} \sqrt{2} & 0 \\ 0 & -\sqrt{2} \end{pmatrix}$. (See Theorem 16.17.)

Theorem 16.22. The minimal polynomial of $T \in A(V)$ is of degree $\leq n$.

*Proof**: We apply induction on $n = \dim V$. If $n=1$, then for any non-zero $v \in V$, $V = Fv$. Thus $T(v) = \alpha v$ for some $\alpha \in F$. This in turn implies that $(T - \alpha I)v = 0$. Since $\{v\}$ is a basis of V , we get $T - \alpha I = 0$. In other words $x - \alpha$ is the minimal polynomial of T . As degree of $x - \alpha$ is 1, the theorem holds for $n=1$.

Now suppose $n > 1$ and assume the validity of the theorem for all vector spaces W of dimension $< n$. Let $\dim V = n$ and let $v (\neq 0) \in V$. Then $n+1$ vectors $v, T(v), T^2(v), \dots, T^n(v)$ are L.D. over F , so there exist α_i 's $\in F$ not all zero such that

$$\alpha_0 v + \alpha_1 T(v) + \alpha_2 T^2(v) + \dots + \alpha_n T^n(v) = 0.$$

Obviously the degree of $f(x) = \alpha_0 + \alpha_1 x + \dots + \alpha_n x^n$ is less than or equal to n , and $[f(T)]v = 0$. If $[f(T)]v = 0 \forall v \in V$, we get that $f(T) = 0$ and so $m(x) \mid f(x)$, where $m(x)$ is the minimal polynomial of T . As a consequence $\deg m(x) \leq n$. So the only case remains to be discussed is the one in which $[f(T)]V \neq (0)$. Let $W = \{v \mid [f(T)]v = 0\}$. W is a non-zero proper subspace of V . Let $\dim W = k$ then $1 \leq k < n$. Let T_1 be the restriction of T to W . Since for any $w \in W$, $[f(T)]w = 0 \Rightarrow f(T)[T(w)] = [Tf(T)](w) = 0 \Rightarrow T(w) \in W, \Rightarrow T_1(w) \in W \Rightarrow T_1 \in A(W)$. By induction hypothesis degree of $m_1(x)$, the minimal polynomial of T_1 is less than or equal to k .

Again consider the quotient space V/W . $\dim V/W = n - k < n$ as $k \geq 1$. Define $T_2 : V/W \rightarrow V/W$ by $T_2(v+W) = T(v) + W \forall v \in V$. Since for $v, v' \in V$, $v+W = v'+W \Rightarrow v-v' \in W \Rightarrow T(v-v') \in W \Rightarrow T(v)+W = T(v')+W$. T_2 is well-defined. It can be easily verified that $T_2 \in A(V/W)$. Again by induction hypothesis, degree of $m_2(x)$, the minimal polynomial of T_2 is less than or equal to $n-k$. Now $m_2(T_2)(V/W) = \bar{0} \Rightarrow m_2(T)V \subseteq W$, so $m_1(T)[m_2(T)V] \subseteq m_1(T)W = m(T_1)W = (0) \Rightarrow [m_1(T)m_2(T)]V = (0) \Rightarrow m(x) \mid m_1(x)m_2(x) \Rightarrow \deg m(x) \leq \deg m_1(x) + \deg m_2(x) \Rightarrow \deg m(x) \leq k + n - k$, i.e., n .

Hence the theorem holds for vector spaces of dimension n too. ■

PROBLEMS

1. Find the minimal polynomials of each of the following matrices over $\mathbb{Q}$ and if any of them is non-singular find also its inverse.

$$\begin{pmatrix} 2 & 1 \\ -1 & 2 \end{pmatrix}; \begin{pmatrix} 3 & -3 \\ -2 & 2 \end{pmatrix}; \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & -2 & 3 \end{pmatrix}; \begin{pmatrix} 2 & 0 & 0 \\ -1 & 3 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

*This proof is due to M.D. Burrow. *The American Mathematical Monthly*, December 1973.

[Hint. Use the analogue of Theorem 16.22 for matrices.]

2. Show that the minimal polynomial of $\begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & -1 & 0 & 3 \end{pmatrix}$

is $x^4 - 3x^3 + x - 1 = 0$.

3. Construct a linear transformation T on a two-dimensional vector space V over $\mathbf{Q}$ whose minimal polynomial is $x^2 + 1$. Let A be the matrix of T relative to some ordered basis of V . Find the eigen values of A regarding A as a matrix over $\mathbf{C}$.

4. Let $A = (a_{ij})$ be an $n \times n$ matrix over F :

(i) Show that any $(\alpha_1, \alpha_2, \dots, \alpha_n) \in F^{(n)}$ is an eigen vector of A if and only if for some $\alpha \in F$, $\alpha(\alpha_1, \alpha_2, \dots, \alpha_n) = (\alpha_1, \alpha_2, \dots, \alpha_n)A$.

(ii) For any eigen value α of A , the set of all eigen vectors of A belonging to α is a subspace of $F^{(n)}$.

5. Determine the eigen values and eigen vectors of each of the matrices in Problem 1.

6. Using Theorem 12.44 prove that given any $n \times n$ matrix A over a field F such that A has n distinct eigen values in F then A is similar to a diagonal matrix over F , i.e., there exists a non-singular matrix B in $M_n(F)$ such that $B^{-1}AB$ is a diagonal matrix.

7. Let $T \in A(V)$ and let $\alpha_1, \alpha_2, \dots, \alpha_k$ be the totality of its distinct eigen values. Suppose that each of α_i has its eigen space of dimension 1. Show that V has an ordered basis $(v_1, v_2, \dots, v_n)$ such that the matrix of T relative to this basis is of the form

$$\left(\begin{array}{cccc|c} \alpha_1 & 0 & \dots & 0 & A \\ 0 & \alpha_2 & \dots & 0 & \\ 0 & 0 & \dots & \alpha_k & \\ \hline & O & & & B \end{array} \right)$$

where A is $k \times n-k$, B is $(n-k) \times (n-k)$ matrices and O is $n-k \times k$ zero matrix.

8. Show that eigen values of $\begin{pmatrix} 3 & 2 & 5 \\ 0 & 10 & 9 \\ 0 & 0 & 6 \end{pmatrix}$ are 3, 10 and 6 or

$\mathbf{Q}$. Determine the eigen space belonging to each of these eigen values. Verify that the direct sum of these eigen spaces is $\mathbf{Q}^3$ itself.

[Hint. Use Corollary 16.16.]

9. Let $T \in A(V)$ be such that $T^2 = T$. Prove that $V = W_1 \oplus W_2$ where W_1 and W_2 are subspaces such that $T(W_1) = (0)$ and $T(w_2) = w_2$ $\forall w_2 \in W_2$.

10. Show that $T \in A(V)$ is singular if and only if and only if there exists $S \in A(V)$, $S \neq 0$ such that $ST = TS = 0$.

11. Prove that $T \in A(V)$ is right invertible if and only if it is left invertible.

In this result true for infinite-dimensional vector spaces?

[Ans. No]

12. V is any vector space over $\mathbf{C}$ and $T \in A(V)$ satisfies $T^3 = T$. Prove that $V = W_1 \oplus W_2 \oplus W_3$ where for each $w_1 \in W_1$, $w_2 \in W_2$, $w_3 \in W_3$, $T(w_1) = 0$, $T(w_2) = w_2$ and $T(w_3) = -w_3$ respectively.

13. If $\text{char } F = 0$, prove that $T \in A(V)$ can be written as a sum of two regular transformations.

14. Definition: A subspace W of V is called *invariant* under any $T \in A(V)$ if $T(W) \subseteq W$.

Let S be a projection on V and $T \in A(V)$. Prove that $S(V)$ is invariant under T if and only if $STS = TS$. Prove that both $S(V)$ and $\text{Ker } S$ are invariant under T if and only if $ST = TS$.

15. Show that a non-zero $T \in A(V)$ is regular if and only if whenever $V = W_1 \oplus W_2$, W_1, W_2 being subspaces of V , then

$$T(V) = T(W_1) \oplus T(W_2).$$

16. Consider $\mathbf{R}^2$. Define $T \in A(\mathbf{R}^2)$ by $T(e_1) = 2e_1$ and $T(e_2) = e_1 + 2e_2$ where $\{e_1, e_2\}$ is standard basis of $\mathbf{R}^2$. Let W_1 be the subspace of $\mathbf{R}^2$ spanned by e_1 .

(a) Prove that W_1 is invariant under T .

(b) Prove that there is no subspace W_2 of $\mathbf{R}^2$ which is invariant under T and also has the property that $\mathbf{R}^2 = W_1 \oplus W_2$.

2. More on Minimal Polynomials

As in the previous section V is a vector space of dimension n over a field F . Let $T (\neq 0) \in F[T]$, we know that $F[T]$ is subring of $A(V)$. Now for any $\alpha \in F$, $f(T) = \alpha_0 I + \alpha_1 T + \dots + \alpha_k T^k \in F[T]$; $\alpha_i \in F$, we have $\alpha f(T) = \alpha \alpha_0 I + \alpha \alpha_1 T + \dots + \alpha \alpha_k T^k \in F[T]$. So that $F[T]$ is a subspace of $A(V)$. As $\dim_F A(V) = n^2$, $\dim_F F[T] \leq n^2$. We have the following.

Lemma. 16.23. If degree of the minimal polynomial $m(x)$ of T is k , then (i) $\{I, T, T^2, \dots, T^{k-1}\}$ forms a basis of $F[T]$ over F ; (ii) dimension of $F[T]$ over F is equal to k .

Proof: (ii) Let $m(x) = x^k - \alpha_1 x^{k-1} - \alpha_2 x^{k-2} - \dots - \alpha_k$. Now $m(T) = 0$ yields $T^k = \alpha_1 T^{k-1} + \alpha_2 T^{k-2} + \dots + \alpha_k I$... (1)

$$\begin{aligned} T^{k+1} &= T^k T = \alpha_1 T^k + \alpha_2 T^{k-1} + \dots + \alpha_k T \\ &= \alpha_1 (\alpha_1 T^{k-1} + \alpha_2 T^{k-2} + \dots + \alpha_k I) + \alpha_2 T^{k-1} + \dots + \alpha_k T \end{aligned} \quad \dots (2)$$

By (2), T^{k+1} is clearly a linear combination of $I, T, T^2, \dots, T^{k-1}$. Again multiplying (2) by T , and substituting value of T^k from

(1), we get that T^{k+2} is also a linear combination of $I, T, T^2, \dots, T^{k-1}$. Proceeding inductively we get that every T^l ($l \geq k$) is a linear combination of $I, T, T^2, \dots, T^{k-1}$. Since every $f(T) = \lambda_0 I + \lambda_1 T + \dots + \lambda_s T^s$ is a linear combination of $I, T, T^2, \dots, T^k$; so if $s > k-1$ substituting the value of each T^l ($l \geq k$) in terms of $I, T, T^2, \dots, T^{k-1}$ we obtain that $f(T)$ is a linear combination of $I, T, T^2, \dots, T^{k-1}$. Hence $F[T]$ as a vector space over F is generated by $I, T, T^2, \dots, T^{k-1}$. If these were L.D. over F then for some β_i ($0 \leq i \leq k-1$) not all zero we would have

$$\beta_0 I + \beta_1 T + \dots + \beta_{k-1} T^{k-1} = 0.$$

Thus T is a root of a non-zero polynomial over F , of degree $< k = \deg m(x)$. This is against the result of Theorem 16.7.

So we get a contradiction. Thus $I, T, T^2, \dots, T^{k-1}$ are L.I. over F . Hence $\{I, T, T^2, \dots, T^{k-1}\}$ is a basis of $F[T]$.

(ii) follows immediately from (i). ■

REMARK. For instance in Example 3, the linear transformation T on $\mathbb{Q}^3$ has $x-2$ as its minimal polynomial. Since degree of $x-2$ is one, we get that $\dim_{\mathbb{Q}} \mathbb{Q}[T] = 1$. But $\mathbb{Q}$ is embeddable in $\mathbb{Q}[T]$, so $\mathbb{Q} \cong \mathbb{Q}[T]$ as rings (hence as fields).

In Example 4, the linear transformation T on $\mathbb{Q}^3$ has x^3-1 as its minimal polynomial. So $\{I, T, T^2\}$ is a basis of $\mathbb{Q}[T]$ over $\mathbb{Q}$. Notice that as x^3-1 is not an irreducible polynomial over $\mathbb{Q}$, the minimal polynomial of a linear transformation need not be irreducible (Cf. Theorem 13.18).

Theorem 16.24. Let $T \in A(V)$ and $m(x)$ be the minimal polynomial of T over F . For any non-zero $v \in V$.

(a) $F[T]v = \{[f(T)](v) \mid f(T) \in F[T]\}$ is a non-zero subspace of V containing v .

(b) There exists a unique non-zero monic polynomial $m_v(x)$ over F such that (i) $[m_v(T)](v) = 0$.

(ii) For any $f(x) \in F[x]$, $[f(T)]v = 0 \Rightarrow m_v(x) \mid f(x)$.

(iii) $m_v(x) \mid m(x)$.

(iv) $\deg m_v(x) = \dim_F F[T]v$. Indeed $\{v, T(v), T^2(v), \dots, T^{t-1}(v)\}$ is a basis of $F[T]v$ where $t = \deg m_v(x)$.

Proof: (a) Clearly $0 \in F[T]v$. Also for any $[f(T)]v, [g(T)]v \in F[T]v$ and $\alpha \in F$; $[f(T)]v - [g(T)]v = [f(T) - g(T)]v \in F[T]v$ and $\alpha[f(T)]v = [\alpha f(T)]v \in F[T]v$. It follows that $F[T]v$ is a subspace of V . As $I \in F[T]$ we get $v = Iv \in F[T]v$. As v is non-zero, $F[T]v$ is a non-zero subspace of V .

(b) Let $W = \{f(x) \mid [f(T)]v = 0\}$. Since $m(T) = 0$, $m(x) \in W$. Thus W is a non-empty subset of $F[x]$. Now for any $f(x), g(x) \in W$ and

$h(x) \in F[x]$, $[f(T)](v)=0$, $[g(T)](v)=0$ yield that $[f(T)-g(T)](v) = [f(T)](v) - [g(T)](v) = 0$ and $[h(T)] [f(T)](v) = h(T) [f(T)](v) = 0$. Consequently $f(x)-g(x) \in W$ and $h(x)f(x) \in W$. Since $F[x]$ is a commutative ring, $f(x)h(x) \in W$. Hence W is a non-zero ideal of $F[x]$. [Recall that $m(x) (\neq 0) \in W$]. Hence it has a unique monic generator, say, $m_v(x)$. (This sort of argument has been used before, see for example Theorem 16.7). Thus (i) follows. As in Theorem 16.7 definition of $m_v(x)$ gives (ii) immediately. Since $[m(T)]v=0$, (ii) $\Rightarrow$ (iii).

Finally let $\deg m_v(x) = t$. Then $m_v(x) = x^t - \lambda_{t-1}x^{t-1} - \lambda_{t-2}x^{t-2} - \dots - \lambda_0$ for some $\lambda_i \in F$. Hence, $[m_v(T)]v=0$ yields $(T^t)v = \lambda_{t-1}T^{t-1}(v) + \lambda_{t-2}T^{t-2}(v) + \dots + \lambda_0 v$. Using similar arguments as in the preceding Lemma, it follows that $\{v, T(v), \dots, T^{t-1}(v)\}$ is a basis of $F[T]v$. Hence $\dim_F F[T]v = t = \deg m_v(x)$. Consequently (iv) is proved. ■

This theorem motivates the following :

Definition 16.25. Let T be any non-zero linear transformation on V . For any non-zero vector $v \in V$, a monic polynomial $m_v(x)$ over F is called the *minimal polynomial of v relative to T* if (i) $[m_v(T)](v)=0$ and for any $f(x) \in F[x]$, $[f(T)]v=0$ yields that $m_v(x)$ divides $f(x)$.

Note : $\deg m_v(x) \geq 1$ for $v \neq 0 \in V$.

Definition 16.26. Let T be a non-zero linear transformation on V . For any $v \in V$ the ideal $W = \{f(x) \in F[x] \mid [f(T)]v=0\}$ of $F[x]$ is called the *annihilator ideal of v with respect to T* .

Clearly $m_v(x)$ is a generator of the annihilator ideal W of v .

Thus we see that the minimal polynomial $m_v(x)$ of v relative to T is a factor of the minimal polynomial $m(x)$ of T and for no non-zero polynomial $f(x)$ of degree $< \deg m_v(x)$, $[f(T)](v)=0$. We consider $m_v(x)$ for each $(v \neq 0) \in V$. Let $m_1(x)$ be their LCM. Since each $m_v(x)$ is a factor of $m(x)$, $m_1(x) \mid m(x)$ and $[m_1(T)]v=0 \forall v \in V$. So that $m_1(T)=0$. Consequently we get $m(x) \mid m_1(x)$. Thus $m(x) = m_1(x)$, the LCM of all $m_v(x)$ (for all $v \neq 0 \in V$).

Note : As for each $v (\neq 0) \in V$, $m_v(x) \mid m(x)$ and $\deg m(x)$ is a positive integer, number of distinct $m_v(x)$'s must be finite. So we are justified in talking of LCM of $m_v(x)$'s.

The above reasoning yields the following.

Corollary 16.27. For any non-zero $T \in A(V)$, the minimal polynomial of T is the LCM of the minimal polynomials relative to T of all the non-zero elements of V . ■

EXAMPLE 9. Consider the linear transformation $T: \mathbb{Q}^1 \rightarrow \mathbb{Q}^2$ whose matrix relative to the standard basis (e_1, e_2) is

$$A = \begin{pmatrix} 0 & 3 \\ 1 & 2 \end{pmatrix}$$

Since $A^2 - 2A - 3I' = 0$ where $I' = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ and as A is not a scalar matrix (i.e. of the type $\alpha I'$ for some $\alpha \in \mathbb{Q}$), A cannot be a root of a polynomial over $\mathbb{Q}$ of degree 1. Hence $x^2 - 2x - 3$ is the minimal polynomial of A and hence of T .

Now $T(e_1) = e_2$, $T(e_2) = 3e_1 + 2e_2$ yield that $T(e_1 + e_2) = T(e_1) + T(e_2) = 3(e_1 + e_2)$. So for $v = e_1 + e_2$, $(T - 3I)(v) = 0$. As $x - 3$ is a monic polynomial of degree 1 and $(T - 3I)(v) = 0$, $x - 3$ itself is the minimal polynomial of v relative to T . It can be similarly seen that the minimal polynomial of $w = -3e_1 + e_2$ relative to T is $x + 1$.

Clearly $x^2 - 2x - 3$ is the LCM of $x + 1$ and $x - 3$.

Consider $v = 2e_1 + e_2$. For any $\alpha \in \mathbb{Q}$,

$$\begin{aligned} (T - \alpha I)(v) &= T(2e_1 + e_2) - \alpha(2e_1 + e_2) = 2T(e_1) + T(e_2) - 2\alpha e_1 - \alpha e_2 \\ &= (4 - \alpha)e_2 + (3 - 2\alpha)e_1 \neq 0, \text{ since for any } \alpha, 4 - \alpha = 0 \\ &\text{and } 3 - 2\alpha = 0 \text{ cannot hold simultaneously.} \end{aligned}$$

Hence no monic polynomial of degree 1 is in the annihilator ideal of v , relative to T . This shows that the minimal polynomial of v must be of degree at least two. Since $(T^2 - 2T - 3I)(v) = 0v = 0$, we get $x^2 - 2x - 3$ is the minimal polynomial of v relative to T . So we see that $\mathbb{Q}^2$ has a vector whose minimal polynomial relative to T is same as that of T . This fact holds for general situation as we shall see below.

Lemma 16.28. Let v, w be two non-zero vectors in V such that their minimal polynomials relative to a non-zero linear transformation T on V are relatively prime to each other. Then the minimal polynomial of $v + w$ is the product of the minimal polynomials of v and w .

Proof: Let $f(x)$ and $g(x)$ be the minimal polynomials of v and w respectively. We are given that $(f(x), g(x)) = 1$. Now $[f(T)]v = 0$, $[g(T)]w = 0$ yield $f(T)g(T)(v + w) = g(T)[f(T)v] + f(T)[g(T)w] = g(T)0 + f(T)0 = 0$. Hence the minimal polynomial $m(x)$ of $v + w$ divides $f(x)g(x)$. As $f(x)$ and $g(x)$ are relatively prime we get that $m(x) = f_1(x)g_1(x)$ for some $f_1(x) \mid f(x)$ and $g_1(x) \mid g(x)$ and $[f_1(x), g_1(x)] = 1$. As $m(x)$ is monic, both $f_1(x)$ and $g_1(x)$ can be taken to be monic polynomials. Suppose $f_1(x) \neq f(x)$, then

$$\deg f_1(x) < \deg f(x) \text{ and hence } [f_1(T)]v \neq 0 \quad \dots (1)$$

$$\begin{aligned} \text{Also } f_1(x)g_1(x) \mid f_1(x)g(x) \Rightarrow 0 &= f_1(T)g(T)(v + w) \\ &= f_1(T)[g(T)](v) \text{ as } [g(T)]w = 0 \quad \dots (2) \end{aligned}$$

As $[f(x), g(x)] = 1$, there exist $A(x), B(x) \in F[x]$ such that $f(x)A(x) + g(x)B(x) = 1$. (Corollary 10.20). Consequently in $F[T]$, $f(T)A(T) + g(T)B(T) = I$.

$$\text{Therefore } v = Iv = [B(T)g(T)](v) \text{ as } [f(T)](v) = 0 \quad \dots (3)$$

Operating $B(T)$ on (2) we get

$$0 = [B(T)f_1(T)g(T)](v) = f_1(T)(Bv) = [f_1(T)]v \text{ by (3). This is a con-}$$

tradition to (1). Hence $\deg f_1(x) = \deg f(x)$. In other words $f_1(x) = f(x)$. Similarly it can be shown that $g_1(x) = g(x)$. Hence $m(x) = f(x)g(x)$. ■

Lemma 16.29. *If non-zero vectors $v_i (i=1, 2, \dots, s)$ in V have respective minimal polynomials $f_i(x)$. ($i=1, 2, \dots, s$) relative to a linear transformation T such that $f_i(x)$ are relatively prime in pairs then $v = v_1 + v_2 + \dots + v_s$ has the product $f_1(x) f_2(x) \dots f_s(x)$ as its minimal polynomial.*

Proof: For $s=2$ the result follows from Lemma 16.28. To apply induction let $s > 2$ and the result holds for $s-1$. Then $w = v_1 + v_2 + \dots + v_{s-1}$ has $g(x) = f_1(x) f_2(x) \dots f_{s-1}(x)$ as its minimal polynomial. As $[f_s(x), f_i(x)] = 1$ for all $1 \leq i < s$, we get $[g(x), f_s(x)] = 1$. Hence by Lemma 16.28, $v = v_1 + v_2 + \dots + v_s = w + v_s$ has $g(x) f_s(x) = f_1(x) f_2(x) \dots f_s(x)$ as its minimal polynomial. Thus the result follows by induction. ■

The following lemma is easy, we leave its proof to the readers.

Lemma 16.30. *If the minimal polynomial of a non-zero $v \in V$ relative to a linear transformation T on V is $m(x) = f(x)g(x)$ where $f(x)$ and $g(x)$ are monic polynomials in $F[x]$, then the minimal polynomial of $f(T)v$ is $g(T)$ relative to T . ■*

Theorem 16.31. *Let T be a non-zero linear transformation on V . Then V has a vector v such that the minimal polynomials of T and of v relative to T are equal.*

Proof: Let $\{v_1, v_2, \dots, v_n\}$ be a basis of V . Let for each $i=1, 2, \dots, n$; $f_i(x)$ be the minimal polynomial of v_i relative to T . Since $F[x]$ is a UFD, we can express these $f_i(x)$'s as products of powers of distinct monic irreducible polynomials.

$$f_i(x) = \Pi_1^{\alpha_{1i}}(x) \Pi_2^{\alpha_{2i}}(x) \dots \Pi_t^{\alpha_{ti}}(x), \quad (i=1, 2, \dots, n)$$

where $\alpha_{ji} \geq 0$ for all i, j and $\Pi_1(x), \Pi_2(x), \dots, \Pi_t(x)$ are distinct irreducible polynomials (we can take same irreducible polynomials for all $f_i(x)$ by allowing zero powers, and $\Pi_j^{\alpha_{ji}}(x)$ stands for $[\Pi_j(x)]^{\alpha_{ji}}$

(cf. Proof of Lemma 10.31). For each $j=1, 2, \dots, t$, let

$$\alpha_j = \max(\alpha_{j1}, \alpha_{j2}, \dots, \alpha_{jn}) \quad \dots (2)$$

Then LCM of $f_1(x), f_2(x), \dots, f_n(x)$ is

$$f(x) = \Pi_1^{\alpha_1}(x) \Pi_2^{\alpha_2}(x) \dots \Pi_t^{\alpha_t}(x)$$

Now $\alpha_1 = \alpha_{1i_1}$, for some i_1 . So

$$f_1(x) = \Pi_1^{\alpha_1}(x) [\Pi_2^{\alpha_{21}}(x) \dots \Pi_t^{\alpha_{t1}}(x)]$$

Hence if $\omega_1 = [\Pi_2^{\alpha_{21}}(T) \dots \Pi_t^{\alpha_{t1}}(T)] v_1$ we have that the minimal

polynomial of ω_1 is $\Pi_1^{\alpha_1}(x)$. (Lemma 16.30). Similarly considering $\alpha_2, \alpha_3, \dots, \alpha_t$ we get vectors $\omega_2, \omega_3, \dots, \omega_t$ whose respective minimal polynomials relative to T are $\Pi_2^{\alpha_2}(x), \dots, \Pi_t^{\alpha_t}(x)$. Then as $\Pi_1^{\alpha_1}(x), \dots, \Pi_t^{\alpha_t}(x)$ are relatively prime in pairs, by Lemma 16.29, $\omega = \omega_1 + \omega_2 + \dots + \omega_t$ has minimal polynomial as

$$f(x) = \Pi_1^{\alpha_1}(x) \Pi_2^{\alpha_2}(x) \dots \Pi_t^{\alpha_t}(x).$$

Let $m(x)$ be the minimal polynomial of T . Let $v \in V$. Then $v = \beta_1 v_1 + \beta_2 v_2 + \dots + \beta_n v_n$; $\beta_i \in F$

$$\text{Thus } [f(T)]v = \beta_1 f(T)v_1 + \dots + \beta_n f(T)v_n = 0, \text{ since } f_i(T)v_i = 0 \text{ and } f_i(x) \mid f(x) \forall i.$$

Hence $f(T)v = 0$. This yields $m(x) \mid f(x)$. Also $f_i(x) \mid m(x)$ for all $1 \leq i \leq t$ (Theorem 16.29) and $f(x)$ is LCM of these $f_i(x)$'s we get $f(x) \mid m(x)$. Hence $f(x) \mid m(x)$ and $m(x) \mid f(x)$. This yields that $f(x) = m(x)$, as both are monic. Hence the minimal polynomials of ω and T are the same, viz., $m(x)$. ■

Theorem 16.22 now follows as a corollary to the above Theorem.

Corollary 16.32. For any non-zero $T \in A(V)$, degree of its minimal polynomial is $\leq n$, the dimension of V .

Proof. Theorem 16.31 gives that V has a vector v such that the minimal polynomial of $f(x)$ of v relative to T is equal to the minimal polynomial of T . Now $F[T]v$ being a subspace of V , $\dim_F F[T]v \leq n = \dim V$. However by Theorem 16.24, $\deg m(x) = \dim_F F[T]v$. Hence $\deg m(x) \leq n$. This proves the corollary. ■

Corollary 16.33. For any non-zero linear transformation T on V , every root in F of the minimal polynomial of T is an eigen value of T .

Proof: Let $m(x)$ be the minimal polynomial of T and α be a root of $m(x)$ in F .

Then $m(x) = (x - \alpha)m_1(x)$; $m_1(x) \in F[x]$. By Theorem 16.31, V has a vector v whose minimal polynomial relative to T is $m(x)$. Then $m_1(T)v \neq 0$ and if $w = m_1(T)v$, we see that $0 = m(T)v = (T - \alpha)w$ yields $T(w) = \alpha w$. Hence α is an eigen value of T . ■

EXAMPLE 10. Consider the polynomial $f(x) = (x^2 + 1)^2 = x^4 + 2x^2 + 1$. We determine linear transformation on $\mathbb{C}^4$ whose minimal polynomial is $x^4 + 2x^2 + 1$. Let (e_1, e_2, e_3, e_4) be an ordered standard basis of $\mathbb{C}^4$. Define a linear transformation T on $\mathbb{C}^4$ such that $T(e_1) = e_2, T(e_2) = e_3, T(e_3) = e_4$ and $T(e_4) = -e_1 - 2e_3$. So $T(e_1) = e_2,$

Theorem 16.35. Let $f(x) = x^m - \alpha_{m-1}x^{m-1} - \dots - \alpha_1x - \alpha_0$ be any monic polynomial over a field F . Then there exists a linear transformation T on a vector space W of dimension $m = \deg f(x)$, such that the minimal polynomial of T is $f(x)$.

Proof: Let $(v_1, v_2, \dots, v_m)$ be an ordered basis of W . Define a linear transformation T on W such that $T(v_1) = v_2, T(v_2) = v_3, \dots, T(v_{m-1}) = v_m, T(v_m) = \alpha_0 v_1 + \alpha_1 v_2 + \dots + \alpha_{m-1} v_m$. Now $\{v_1, T(v_1), T^2(v_1), \dots, T^{m-1}(v_1)\} = \{v_1, v_2, \dots, v_m\}$ is a L.I. set, so by Theorem 16.24, the degree of the minimal polynomial of v_1 relative to T is at least m and hence m , as $\dim W = m$; also degree of the minimal polynomial of any vector or of a linear transformation cannot exceed the dimension of underlying vector space (Corollary 16.32). Now the defining relations of T give $T^m(v_1) = \alpha_0 v_1 + \alpha_1 T(v_1) + \alpha_2 T^2(v_1) + \dots + \alpha_{m-1} T^{m-1}(v_1)$ i.e. $(T^m - \alpha_{m-1} T^{m-1} - \alpha_{m-2} T^{m-2} - \dots - \alpha_1 T - \alpha_0 I)v_1 = 0$. Hence $f(x) = x^m - \alpha_{m-1}x^{m-1} - \alpha_{m-2}x^{m-2} - \dots - \alpha_0$ is the minimal polynomial of v_1 relative to T . Again by appealing to Theorem 16.24 (b) (iii) and Corollary 16.32, we get that $f(x)$ is the minimal polynomial of T . ■

Theorem 16.36. If the minimal polynomial $f(x)$ of a linear transformation T on V is of degree equal to the dimension of V , then V has an ordered basis such that matrix of T relative to that basis is $C(f(x))$.

Proof: Let $f(x) = x^k - \alpha_{k-1}x^{k-1} - \dots - \alpha_0$ be the minimal polynomial of $T \in A(V)$ where $\dim V = k$. By Theorem 16.31 there exists a vector $v_1 (\neq 0) \in V$ such that $f(x)$ is minimal polynomial of v_1 relative to T . Hence $\{v_1, T(v_1), T^2(v_1), \dots, T^{k-1}(v_1)\}$ is a basis of V (Theorem 16.24). Put $v_2 = T(v_1), v_3 = T^2(v_1) = T(v_2), \dots, v_k = T^{k-1}(v_1) = T(v_{k-1})$. Now $(v_1, v_2, \dots, v_k)$ is an ordered basis of V such that $T(v_1) = v_2, T(v_2) = v_3, \dots, T(v_{k-1}) = v_k$ and $T(v_k) = T^k(v_1) = \alpha_0 v_1 + \alpha_1 T(v_1) + \dots + \alpha_{k-1} T^{k-1}(v_1) = f(T)v_1 = 0$.

This gives that $T(v_k) = \alpha_0 v_1 + \alpha_1 v_2 + \dots + \alpha_{k-1} v_k$. The matrix of T relative to the ordered basis $(v_1, v_2, \dots, v_k)$ is

$$\begin{pmatrix} 0 & 0 & \dots & 0 & \alpha_0 \\ 1 & 0 & \dots & 0 & \alpha_1 \\ 0 & 1 & \dots & 0 & \alpha_2 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 0 & \alpha_{k-2} \\ 0 & 0 & \dots & 1 & \alpha_{k-1} \end{pmatrix} = C(f(x)). \blacksquare$$

Corollary 16.37. Let A be an $n \times n$ matrix over a field F . If the minimal polynomial of $f(x)$ of A is of degree n , then there exists a non-singular matrix P over F such that $P^{-1}AP$ is $C(f(x))$.

Proof: Let V be an n dimensional vector space over F and $\mathbf{B} = (v_1, v_2, \dots, v_n)$ be an ordered basis of V . Let T be a L.T. on V whose matrix relative to $\mathbf{B}$ is A . Then the minimal polynomial of T is also

$f(x)$, [Definition 16.9 (a)]. Since $\deg f(x)$ is n , there exists an ordered basis $\mathbf{B}' = (w_1, w_2, \dots, w_n)$ of V such that matrix of T relative to $\mathbf{B}'$ is $A' = C(f(x))$. If P is the matrix of $(v_1, v_2, \dots, v_n)$ relative to $(w_1, w_2, \dots, w_n)$ then $A' = P^{-1}AP$ (Corollary 12.40). Hence $P^{-1}AP = C(f(x))$. ■

Definition 16.38. A linear transformation T on a vector space V is said to be *cyclic* if there exists $v \in V$ such that $V = F[T]v$; in such a situation V is said to be a *cyclic space relative to T* .

Lemma 16.39. A vector space V is cyclic relative to a linear transformation T on V if and only if $\dim_F V$ is same as the degree of the minimal polynomial of T .

Proof: If V is cyclic relative to T , then there exists $v \in V$ such that $V = F[T]v$. By Theorem 16.24, $\dim_F F[T]v = \deg m_v(x)$ where $m_v(x)$ is the minimal polynomial of v relative to T . So $\deg m_v(x) = n = \dim V$. Then $m_v(x)$ is itself the minimal polynomial of T (Theorem 16.24 and Corollary 16.32).

Conversely let the degree of $m(x)$, the minimal polynomial of T , be n where $n = \dim V$. Now V has a vector v whose minimal polynomial relative to T is $m(x)$ and hence $\dim F[T]v = n$ [Theorem 16.24 (b) (iv)]. Consequently $V = F[T]v$. This proves that V is cyclic relative to T . ■

Definition 16.40. Let T be a cyclic linear transformation on V and $f(x) = x^n - \alpha_{n-1}x^{n-1} - \dots - \alpha_1x - \alpha_0$ be its minimal polynomial. Then the companion matrix of $f(x)$ viz,

$$A = \begin{pmatrix} 0 & 0 & \dots & 0 & \alpha_0 \\ 1 & 0 & \dots & 0 & \alpha_1 \\ 0 & 1 & \dots & 0 & \alpha_2 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 0 & \alpha_{n-2} \\ 0 & 0 & \dots & 1 & \alpha_{n-1} \end{pmatrix}$$

is called *Jordan Matrix* of T .

Theorem 16.41. Let $f(x) = (q(x))^k$, be a monic polynomial over F of degree n with $q(x)$ being any monic polynomial of degree m (i.e. $n = mk$). If T is cyclic linear transformation on V such that $f(x)$ is the minimal polynomial of T then, relative to some ordered basis of V , the matrix of T is of the form

$$A = \begin{pmatrix} C & N & & \\ & C & N & \\ & & \ddots & N \\ & & & C \end{pmatrix}$$

where $C = C(q(x))$ and N is the $m \times m$ matrix

$$\begin{pmatrix} 0 & 0 & \dots & 1 \\ 0 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 0 \end{pmatrix}$$

Proof: Now for some $v \in V$, $V = F[T]v$.

Define

$$\left. \begin{aligned} v_1 &= [q(T)]^{k-1}v, v_2 = T[q(T)]^{k-1}v, \dots, v_m = T^{m-1}[q(T)]^{k-1}v, \\ v_{m+1} &= [q(T)]^{k-2}v, v_{m+2} = T[q(T)]^{k-2}v, \dots, v_{2m} = T^{m-1}[q(T)]^{k-2}v, \\ v_{2m+1} &= [q(T)]^{k-3}v, v_{2m+2} = T[q(T)]^{k-3}v, \dots, v_{3m} = T^{m-1}[q(T)]^{k-3}v, \\ &\dots \dots \dots \end{aligned} \right\} \dots (1)$$

$$v_{(k-1)m+1} = v, v_{(k-1)m+2} = T(v), \dots, v_{km} = T^{m-1}(v).$$

Thus each of the v_i 's is of the form $g_i(T)v$ with $g_i(x) \neq 0$ and $\deg g_i(x) < \deg f(x) = n$ and the minimal polynomials associated with these v_i 's are of degree $< n$. Now for some $\alpha_i \in F$, $\sum_i \alpha_i v_i = 0$ gives

$\sum \alpha_i g_i(T)v = 0$. If some $\alpha_j \neq 0$, the polynomial $g(x) = \sum \alpha_i g_i(x)$ is non-zero and is of degree $< \deg f(x)$ and $[g(T)]v = 0$. This contradicts the fact that $f(x)$ is the minimal polynomial of v (see Theorem 16.24). Hence the n vectors $v_1, v_2, \dots, v_{km} (= v_n)$ are L.I.

$$\text{Let } q(x) = x^m - \alpha_{m-1}x^{m-1} - \dots - \alpha_1x - \alpha_0. \dots (2)$$

The definitions of v_i 's given above yield the following:

$$T(v_1) = v_2$$

$$T(v_2) = v_3$$

$$\dots \dots \dots$$

$$\dots \dots \dots$$

$$T(v_{m-1}) = v_m$$

$$T(v_m) = T^m[q(T)]^{k-1}v$$

$$= [T^m - q(T)][q(T)]^{k-1}v, \text{ since } [q(T)]^k v = 0$$

$$= (\alpha_0 I + \alpha_1 T + \dots + \alpha_{m-1} T^{m-1})[q(T)]^{k-1}v$$

$$= \alpha_0 [q(T)]^{k-1}v + \alpha_1 T[q(T)]^{k-1}v + \dots + \alpha_{m-1} T^{m-1}[q(T)]^{k-1}v$$

This implies that

$$T(v_m) = \alpha_0 v_1 + \alpha_1 v_2 + \dots + \alpha_{m-1} v_m$$

$$T(v_{m+1}) = v_{m+2}$$

$$T(v_{m+2}) = v_{m+3}$$

$$\dots \dots \dots$$

$$T(v_{2m-1}) = v_{2m}$$

$$T(v_{2m}) = T^m[q(T)]^{k-2}v$$

$$= [q(T)]^{k-1}v + [T^m - q(T)][q(T)]^{k-2}v$$

$$= v_1 + (\alpha_0 I + \alpha_1 T + \dots + \alpha_{m-1} T^{m-1})[q(T)]^{k-2}v$$

This gives that

$$T(v_{2m}) = v_1 + \alpha_0 v_{m+1} + \alpha_1 v_{m+2} + \dots + \alpha_{m-1} v_{2m}.$$

Continuing in this manner, we get

$$T(v_{2m+1}) = v_{2m+2}$$

$$T(v_{2m+2}) = v_{2m+3}$$

$$T(v_{3m-1}) = v_{3m}$$

$$T(v_{3m}) = v_{m+1} + \alpha_0 v_{2m+1} + \dots + \alpha_{n-1} v_{3m}$$

$$T(v_{(k-1)m+1}) = v_{(k-1)m+2}$$

$$T(v_{km-1}) = v_{km}$$

$$T(v_{km}) = v_{(k-2)m+1} + \alpha_0 v_{(k-1)m+1} + \alpha_1 v_{(k-1)m+2} + \dots + \alpha_{m-1} v_{km}$$

Using the fact that if $T(v_i) = \sum_{j=1}^{n=km} \alpha_{ji} v_j$, then in the i th column of matrix (α_{ji}) , the j th entry is the coefficient of v_j , all the above equations yield that the matrix of T relative to $(v_1, v_2, \dots, v_{km})$ is

	$\overbrace{\hspace{2cm}}^{mth \text{ column}}$	$\overbrace{\hspace{2cm}}^{2mth \text{ column}}$	$\overbrace{\hspace{2cm}}^{kmth \text{ column}}$
$mth \text{ row}$	$\begin{matrix} 0 & 0 & \dots & 0 & \alpha_0 \\ 1 & 0 & \dots & 0 & \alpha_1 \\ 0 & 1 & \dots & 0 & \alpha_2 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 & \alpha_{m-1} \end{matrix}$	$\begin{matrix} 0 & 0 & \dots & 0 & 1 \\ 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 0 & 0 \end{matrix}$	$\begin{matrix} \dots & 0 & 0 & \dots & \dots & 0 \\ \dots & 0 & 0 & \dots & \dots & 0 \\ \dots & 0 & 0 & \dots & \dots & 0 \\ \dots & 0 & 0 & \dots & \dots & 0 \\ \dots & 0 & 0 & \dots & \dots & 0 \end{matrix}$
$2mth \text{ row}$	$\begin{matrix} 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 0 & 0 \end{matrix}$	$\begin{matrix} 0 & 0 & \dots & 0 & \alpha_0 \\ 1 & 0 & \dots & 0 & \alpha_1 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 & \alpha_{m-1} \end{matrix}$	$\begin{matrix} \dots & 0 & 0 & \dots & \dots & 0 \\ \dots & 0 & 0 & \dots & \dots & 0 \\ \dots & 0 & 0 & \dots & \dots & 0 \\ \dots & 0 & 0 & \dots & \dots & 0 \end{matrix}$
$kmth \text{ row}$	$\begin{matrix} 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 0 & 0 \end{matrix}$	$\begin{matrix} 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 0 & 0 \end{matrix}$	$\begin{matrix} 0 & 0 & \dots & 0 & \alpha_n \\ 1 & 0 & \dots & 0 & \alpha_1 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 0 & \alpha_{m-2} \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 & \alpha_{m-1} \end{matrix}$

$$= \begin{pmatrix} C & N & O & O & \dots & O & O \\ O & C & N & O & \dots & O & O \\ O & O & C & N & \dots & O & O \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ O & O & O & O & \dots & C & N \\ O & O & O & O & \dots & O & C \end{pmatrix}$$

where C is $m \times m$ matrix $\begin{pmatrix} 0 & 0 & \dots & 0 & \alpha_0 \\ 1 & 0 & \dots & 0 & \alpha_1 \\ 0 & 1 & \dots & 0 & \alpha_2 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 & \alpha_{m-1} \end{pmatrix} = C(q(x))$ and N

is the $m \times m$ matrix $\begin{pmatrix} 0 & 0 & \dots & 0 & 1 \\ 0 & 0 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 0 & 0 \end{pmatrix}$.

This proves the theorem. ■

Definition 16.42. Let $q(x)$ be a monic polynomial of degree, say m , over a field F . For any positive integer k , the $km \times km$ matrix

$$\begin{pmatrix} C & N & O & \dots & O & O \\ O & C & N & \dots & O & O \\ \dots & \dots & \dots & \dots & \dots & \dots \\ O & O & O & \dots & C & N \\ O & O & O & \dots & O & C \end{pmatrix}$$

where C is the companion matrix of $q(x)$ and N is the $m \times m$ matrix

$$\begin{pmatrix} 0 & 0 & \dots & 0 & 1 \\ 0 & 0 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & \dots & 0 & 0 \end{pmatrix}$$

is called the *Jordan matrix* of $f(x) = (q(x))^k$ relative to $q(x)$.

EXAMPLE 14. Consider $f(x) = (x^2 + 1)^3$, $f(x)$ is of degree 6 over $\mathbb{Q}$. Here $q(x) = x^2 + 1$. The companion matrix of $q(x)$ is

$$C = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

We have a linear transformation T over $\mathbb{Q}^6$ whose minimal polynomial is $f(x)$ (Theorem 16.35) and there exists an ordered basis $\mathbf{B} = (u_1, u_2, u_3, u_4, u_5, u_6)$ of $\mathbb{Q}^6$ such that the matrix of T relative to this basis is the companion matrix of $f(x) = (x^2 + 1)^3 = x^6 + 3x^4 + 3x^2 + 1$

i.e.

$$A = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & -1 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & -3 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & -3 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

At the same time, Theorem 16.41 yields that there exists an ordered basis $\mathbf{B}' = (v_1, v_2, v_3, v_4, v_5, v_6)$ of $\mathbb{Q}^6$ such that the matrix of T relative to $\mathbf{B}'$ is $A' = \begin{pmatrix} C & N & O \\ O & C & N \\ O & O & C \end{pmatrix}$ where $N = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$

$$\text{Thus } A' = \begin{pmatrix} 0 & -1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 \end{pmatrix}$$

So if P is the matrix of B' relative to B then $A' = P^{-1}AP$. Hence A' and A are similar matrices, A' is the Jordan matrix of $f(x)$ relative to $x^2 + 1$.

EXAMPLE 15. Let $\alpha \in F$ and $f(x) = (x - \alpha)^n$.

Now $f(x) = x^n - {}^nC_1 x^{n-1} \alpha + {}^nC_2 x^{n-2} \alpha^2 - \dots + (-1)^n \alpha^n$ where for

$$1 \leq r \leq n, {}^nC_r = \frac{n(n-1)\dots(n-r+1)}{1.2.3\dots r}.$$

So the companion matrix of $f(x)$ is

$$A = \begin{pmatrix} 0 & 0 & 0 & \dots & 0 & (-1)^{n+1} \alpha^n \\ 1 & 0 & 0 & \dots & 0 & (-1)^n {}^nC_1 \alpha^{n-1} \\ 0 & 1 & 0 & \dots & 0 & (-1)^{n-1} {}^nC_2 \alpha^{n-2} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 1 & {}^nC_{n-1} \alpha \end{pmatrix}$$

At the same time the companion matrix of $q(x) = x - \alpha$ is (α) . If $N = (1)$, then the $n \times n$ matrix

$$A' = \begin{pmatrix} C & N & 0 & \dots & 0 \\ 0 & C & N & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & C & N \\ 0 & 0 & \dots & 0 & C \end{pmatrix} = \begin{pmatrix} \alpha & 1 & 0 & \dots & 0 \\ 0 & \alpha & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & \alpha \\ 0 & 0 & 0 & \dots & \alpha \end{pmatrix}$$

Since A and A' are the matrices of same linear transformation on $F^{(n)}$, A and A' are similar, i.e., there exists a non-singular matrix $P \in M_n(F)$ such that $A' = P^{-1}AP$ (Corollary 12.40).

REMARK. Example 15 is to be looked into more carefully. What we actually see is that given $\alpha \in F$, the $n \times n$ matrix

$$A = \begin{pmatrix} \alpha & 1 & 0 & 0 & \dots & 0 \\ 0 & \alpha & 1 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 & \dots & \alpha \\ 0 & 0 & 0 & 0 & \dots & \alpha \end{pmatrix}$$

is such that $f(x) = (x - \alpha)^n$ is the minimal polynomial of A (Theorems 16.35 and 16.41). Since every eigen value of A is a root of its minimal polynomial and $f(x)$ has α as its only root, we conclude that α is the only eigen value of A . This matrix A , which is a special case of that given in Definition 16.42, is called by many authors, a *Jordan Block*.

Thus we can immediately, say that the matrices

$$\begin{pmatrix} 2 & 1 \\ 0 & 2 \end{pmatrix}, \begin{pmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \\ 0 & 0 & -1 \end{pmatrix}, \begin{pmatrix} 5 & 1 & 0 & 0 \\ 0 & 5 & 1 & 0 \\ 0 & 0 & 5 & 1 \\ 0 & 0 & 0 & 5 \end{pmatrix}$$

have respectively 2, -1, 5 as their only eigen values. So if we can show that a given matrix is similar to a Jordan Block; once we know

the latter, it is very easy to determine its eigen values (there happens to be only one).

Definition 16.43. Let T be a linear transformation on a vector space V . A subspace W of V is said to be T -invariant subspace of V if for every $x \in W$, $T(x) \in W$, [i.e., $T(W) \subseteq W$].

REMARK. For any $T \in A(V)$; V , (0) and $T(V)$ are always T -invariant subspaces of V .

Given a T -invariant subspace W of V , we can define a linear transformation $T' : W \rightarrow W$ such that $T'(x) = T(x)$ for every $x \in W$. T' is said to be induced by T .

Now we have the following.

Lemma 16.44. Let $V = V_1 \oplus V_2$ and T be a linear transformation from V to V such that V_1 and V_2 are both T -invariant subspaces. If T_i ($i=1, 2$) is the linear transformation on V_i ($i=1, 2$) induced by T , then the minimal polynomial $f(x)$ of T is the LCM of the minimal polynomials of T_1 and T_2 .

Proof: Let $f_i(x)$ be the minimal polynomial of T_i . Since $f(T) = 0$ and T_i is the restriction of T to V_i , we also have $f(T_i) = 0$. Hence $f_i(x) \mid f(x)$ (Theorem 16.7). Therefore the LCM $g(x)$ of these $f_i(x)$'s also divides $f(x)$.

Now for any $x_1 \in V_1$, $g(T_1)x_1 = g(T)x_1 = 0$, as $g(T_1) = 0$ [note that $f_i(x) \mid g(x)$]. Similarly $g(T)x_2 = 0$ for every $x_2 \in V_2$. Hence $[g(T)]V_1 = (0)$ and $[g(T)]V_2 = (0)$. This yields that $g(T)V = (0)$ as $V = V_1 \oplus V_2$. Hence $g(T) = 0$. Consequently $f(x) \mid g(x)$. Thus $f(x)$ is an associate of $g(x)$. As a result of this $f(x)$ is LCM of $f_1(x)$ and $f_2(x)$. ■

Theorem 16.45. Let T be a cyclic linear transformation on a vector space V and $f(x)$ be its minimal polynomial. If $f(x) = [f_1(x)^{\alpha_1} f_2(x)^{\alpha_2} \dots f_t(x)^{\alpha_t}]$ is the decomposition of $f(x)$ into product of powers of distinct irreducible monic polynomials $f_i(x)$'s over F . Then $V = V_1 \oplus V_2 \oplus \dots \oplus V_t$ for some T -invariant subspaces V_i of V such that the restriction T_i of T to V_i is a cyclic linear transformation on V_i with $[f_i(x)]^{\alpha_i}$ as its minimal polynomial.

Proof: Since T is a cyclic transformation on V , $V = F[T]v$ for some $v \in V$ and $f(x)$ is the minimal polynomial of v relative to T . Define $v_1 = [f_2(T)]^{\alpha_2} \dots [f_t(T)]^{\alpha_t} v$, $v_2 = [f_1(T)]^{\alpha_1} [f_3(T)]^{\alpha_3} \dots [f_t(T)]^{\alpha_t} v$, $\dots$, $v_t = [f_1(T)]^{\alpha_1} [f_2(T)]^{\alpha_2} \dots [f_{t-1}(T)]^{\alpha_{t-1}} v$.

Then the minimal polynomial of v_1 is $[f_1(T)]^{\alpha_1}$ (Lemma 16.30). In general the minimal polynomial of v_i is $[f_i(x)]^{\alpha_i}$. Thus by Lemma 16.29, the minimal polynomial of

$$w = v_1 + v_2 + \dots + v_t \quad \dots (1)$$

is $[f_1(x)]^{\alpha_1} [f_2(x)]^{\alpha_2} \dots [f_t(x)]^{\alpha_t} = f(x)$.

Hence $\dim V = \deg f(x) = \dim F[T]_w$ yields that

$$V = F[T]w \quad \dots(2)$$

Let $V_i = F[T]v_i$ ($1 \leq i \leq t$). Since $TF[T]v_i \subseteq F[T]v_i$, we get that V_i is T -invariant space. So if T_i is the restriction of T to V_i , we get $V_i = F[T_i]v_i$, a cyclic space relative to T_i . Since $[f_i(x)]^{\alpha_i}$ is the minimal polynomial of v_i relative to T , we also get $[f_i(x)]^{\alpha_i}$ is the minimal polynomial of v_i relative to T_i . Hence T_i also has $[f_i(x)]^{\alpha_i}$ as its minimal polynomial.

Since $w = v_1 + v_2 + \dots + v_t$ yields $V = F[T]w \subseteq F[T]v_1 + \dots + F[T]v_t$, we get

$$V = V_1 + V_2 + \dots + V_t \quad \dots(3)$$

We claim that this sum is direct.

$$\text{Let } g_1(T)v_1 + g_2(T)v_2 + \dots + g_t(T)v_t = 0 \quad \dots(4)$$

$$\text{for some } g_i(T)v_i \in V_i = F[T]v_i. \text{ Then } [f_2(T)]^{\alpha_2} \dots [f_t(T)]^{\alpha_t} [g_1(T)v_1 + \dots + g_t(T)v_t] = 0 \quad \dots(5)$$

Since $[f_i(T)]^{\alpha_i}v_i = 0$ for all i , (5) yields that

$$g_1(T)[f_2(T)]^{\alpha_2} \dots [f_t(T)]^{\alpha_t}v_1 = 0.$$

Then $[f_1(x)]^{\alpha_1} \mid g_1(x)[f_2(x)]^{\alpha_2} \dots [f_t(x)]^{\alpha_t}$ (Theorem 16.24). Consequently $[f_1(x)]^{\alpha_1} \mid g_1(x)$, as $[f_1(x)]^{\alpha_1}$ is relatively prime to each of $[f_i(x)]^{\alpha_i}$, ($i \geq 2$). Hence $g_1(T)v_1 = 0$. Similarly $g_2(T)v_2 = 0, \dots, g_t(T)v_t = 0$. Thus in (4) every term on the left is zero. Hence $V = V_1 \oplus V_2 \oplus \dots \oplus V_t$. This proves the theorem. ■

Theorem 16.46. Let T be a cyclic linear transformation on a vector space V and $f(x) = [f_1(x)]^{\alpha_1} [f_2(x)]^{\alpha_2} \dots [f_t(x)]^{\alpha_t}$ be the minimal polynomial of T , where $f_i(x)$'s are distinct irreducible monic polynomials over F and $\alpha_i \geq 1$. Then there exists an ordered basis $\mathbf{B}$ of V such that the matrix of T relative to $\mathbf{B}$ is

$$\begin{pmatrix} J_1 & & \\ & J_2 & \\ & & \ddots \\ & & & J_t \end{pmatrix}$$

where each J_i ($i = 1, 2, \dots, t$) is the Jordan matrix of $[f_i(x)]^{\alpha_i}$ relative to $f_i(x)$, (Definition 16.42).

Proof: We can write $\dots(1)$

$$V = V_1 \oplus V_2 \oplus \dots \oplus V_t$$

for some T -invariant subspaces V_i of V such that V_i is cyclic relative to T_i , the restriction of T to V_i and $[f_i(x)]^{\alpha_i}$ is the minimal polynomial of T_i (Theorem 16.45). We can find an ordered basis $\mathbf{B}_i = (w_1^i, w_2^i, \dots, w_{n_i}^i)$ of V_i such that the matrix of T_i relative to $\mathbf{B}_i$ is the Jordan matrix J_i of $[f_i(x)]^{\alpha_i}$ relative to $f_i(x)$, (Theorem 16.41 and Definition 16.42). Now

$\mathbf{B} = (w^1_1, w^1_2, \dots, w^1_{n_1}, w^2_1, w^2_2, \dots, w^2_{n_2}, \dots, w^t_1, w^t_2, \dots, w^t_{n_t})$ is an ordered basis of V . Let $J_i = (b^i_{lm})$. Then

$$T(w^i_m) = T_i(w^i_m) = \sum_{l=1}^{n_i} b^i_{lm} w^i_l \quad \forall i, m \quad \dots (2)$$

Observe that $T(w^i_m)$ is expressible as a linear combination of $w^i_1, w^i_2, \dots, w^i_{n_i}$ and the coefficients b^i_{lm} are coming from J_i . Keeping this in mind one can easily convince oneself that the matrix of T relative to $\mathbf{B}$ is

$$\begin{pmatrix} J_1 & & \\ & J_2 & \\ & & \ddots \\ & & & J_t \end{pmatrix} \blacksquare$$

Following is immediate.

Corollary 16.47. Let $A \in M_n(F)$ be such that its minimal polynomial $f(x)$ is of degree n . If $f(x) = [f_1(x)^{\alpha_1} \dots [f_t(x)]^{\alpha_t}]$ where $f_i(x)$'s are irreducible monic polynomials over F and $\alpha_i \geq 1$. Then there exists a non-singular matrix P such that

$$P^{-1}AP = \begin{pmatrix} J_1 & & \\ & J_2 & \\ & & \ddots \\ & & & J_t \end{pmatrix}$$

where J_i is the Jordan matrix of $[f_i(x)]^{\alpha_i}$ relative to $f_i(x)$. $\blacksquare$

We know from the Fundamental Theorem of Algebra (see Appendix) that any polynomial of degree n over $\mathbf{C}$, the complex number field has all its n roots in $\mathbf{C}$. So if $f(x)$ is a monic polynomial over $\mathbf{C}$ of degree n , we can write $f(x) = (x-a_1)^{\alpha_1}(x-a_2)^{\alpha_2} \dots (x-a_t)^{\alpha_t}$ where $a_1, a_2, \dots, a_t$ are distinct roots of $f(x)$ and α_i are positive integers. As seen before, the Jordan matrix J_i of $(x-a_i)^{\alpha_i}$ relative to $x-a_i$ is the Jordan Block

$$\begin{pmatrix} a_i & 1 & 0 & \dots & 0 \\ 0 & a_i & 1 & \dots & 0 \\ 0 & 0 & a_i & \dots & 0 \\ 0 & 0 & 0 & \dots & 1 \\ 0 & 0 & 0 & \dots & a_i \end{pmatrix}$$

Thus Theorem 16.46 yields the following.

Corollary 16.48. Let $A \in M_n(\mathbf{C})$ be such that its minimal polynomial $f(x)$ is of degree n . If $a_1, a_2, \dots, a_t$ are the distinct roots of $f(x)$, of multiplicities $\alpha_1, \alpha_2, \dots, \alpha_t$ respectively, then there exists a non-singular matrix P in $M_n(\mathbf{C})$ such that

$$P^{-1}AP = \begin{pmatrix} J_1 & & \\ & J_2 & \\ & & \ddots \\ & & & J_t \end{pmatrix}$$

where each J_i is the $\alpha_i \times \alpha_i$ Jordan Block

$$\begin{pmatrix} a_i & 1 & 0 & \dots & 0 \\ 0 & a_i & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 1 \\ 0 & 0 & 0 & \dots & a_i \end{pmatrix} \quad \blacksquare$$

Definition 16.49. Let T be a cyclic linear transformation on a vector space V_F and $f(x) = [f_1(x)]^{\alpha_1} \dots [f_t(x)]^{\alpha_t}$, where $f_i(x)$'s are distinct irreducible monic polynomials over F and $\alpha_i \geq 1$, be the minimal polynomial of T . Then the matrix

$$\begin{pmatrix} J_1 & & \\ & J_2 & \\ & & \ddots \\ & & & J_t \end{pmatrix}$$

where J_i is the Jordan matrix of $[f_i(x)]^{\alpha_i}$ relative to $f_i(x)$, is called *Classical Canonical matrix* or *Jordan Canonical form* of T ; the above matrix is also called *Classical Canonical matrix of the polynomial $f(x)$* .

Definition 16.50. Let T be a linear transformation on a vector space V and $V_1, V_2, \dots, V_t$ be T -invariant subspaces of V such that $V = V_1 \oplus V_2 \oplus \dots \oplus V_t$. If T_i is the restriction of T to V_i , then T is called a *direct sum* of $T_1, T_2, \dots, T_t$ and we write $T = T_1 \oplus T_2 \oplus \dots \oplus T_t$.

If the A_i is the matrix of T_i relative to some ordered basis $\mathbf{B}_i$ of V_i , and $\mathbf{B}$ is the ordered basis of V , which is the union of all the $\mathbf{B}_i$'s such that in the ordering, the members of $\mathbf{B}_1$ come first those of $\mathbf{B}_2$ come next, those of $\mathbf{B}_3$ come next and so on. It can be easily seen that the matrix A of T relative to $\mathbf{B}$ is

$$\begin{pmatrix} A_1 & & \\ & A_2 & \\ & & \ddots \\ & & & A_t \end{pmatrix}$$

We say that A is *direct sum* of $A_1, A_2, \dots, A_t$ and write $A = A_1 \oplus A_2 \oplus \dots \oplus A_t$.

EXAMPLE 16. Let T be a linear transformation on $\mathbb{Q}^3$ whose minimal polynomial is $f(x) = (x^2 + 1)(x + 1)$. $x^2 + 1$ and $x + 1$ are distinct monic irreducible factors of $f(x)$. The respective companion matrices of these polynomials are $J_1 = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ and $J_2 = (-1)$. So

minimal polynomial of A (i.e. the polynomial of which A is the companion matrix).

7. Let A_1, A_2 be any two matrices over a field F . Show that the matrices $\begin{pmatrix} A_1 & O \\ O & A_2 \end{pmatrix}$ and $\begin{pmatrix} A_2 & O \\ O & A_1 \end{pmatrix}$ are similar.

[Hint. Let V_1 be a vector space over F having an ordered basis $\mathbf{B}_1 = (u_1, u_2, \dots, u_l)$ such that A_1 is a matrix of a linear transformation T_1 on V_1 relative to the basis $\mathbf{B}_1$. Let A_2 be the matrix of a linear transformation T_2 on a vector space V_2 over F , relative to an ordered basis $\mathbf{B}_2 = (v_1, v_2, \dots, v_m)$. Consider $V = V_1 \oplus V_2$ and $T = T_1 \oplus T_2$. Then the matrix of T relative to the ordered basis $\mathbf{B} = (u_1, u_2, \dots, u_l, v_1, v_2, \dots, v_m)$ is $\begin{pmatrix} A_1 & O \\ O & A_2 \end{pmatrix}$ and relative to the ordered basis

$$\mathbf{B}' = (v_1, v_2, \dots, v_m, u_1, u_2, \dots, u_l) \text{ is } \begin{pmatrix} A_2 & O \\ O & A_1 \end{pmatrix}.$$

8. For any k matrices $C_1, C_2, \dots, C_k$ and for any permutation σ of $\{1, 2, 3, \dots, k\}$, show that the matrices

$$\begin{pmatrix} C_1 & & \\ & C_2 & \\ & & \ddots \\ & & & C_k \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} C_{\sigma(1)} & & \\ & C_{\sigma(2)} & \\ & & \ddots \\ & & & C_{\sigma(k)} \end{pmatrix}$$

are similar.

9. A linear transformation T on a vector space V is said to be *nilpotent* if $T^k = 0$ for some positive integer k . Similar definition can be given for a matrix. Show that the matrices

$$A_2 = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \quad A_3 = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$$

$$A_4 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \text{ are nilpotent.}$$

Indeed verify:

(i) $A_2^2 = O, A_2 \neq O.$

(ii) $A_3^3 = O$ but $A_3^2 \neq O.$

(iii) $A_4^4 = O$ but $A_4^3 \neq O.$

(b) If T is a non-zero nilpotent linear transformation on V such that $T^k = 0$ but $T^{k-1} \neq 0$, then k is called the *index of nilpotence* of T . Similar definition can be given for a matrix.

If $A_1, A_2, A_3, \dots, A_t$ are nilpotent matrices with indices of nilpotence $k_1, k_2, \dots, k_t$ respectively, then prove that the matrix

$$A = \begin{pmatrix} A_1 & & \\ & A_2 & \\ & & \ddots \\ & & & A_t \end{pmatrix} \text{ is nilpotent}$$

and its index of nilpotence is $k = \max(k_i)$, ($1 \leq i \leq t$).

(c) Show that a nilpotent linear transformation does not admit any non-zero eigen value.

[Hint. $T(v) = \alpha v \Rightarrow T^k(v) = \alpha^k v$ for all $k \geq 1$].

4. Some Decomposition Theorems

Theorem 16.51. Let $T \in A(V)$ be such that its minimal polynomial is $f(x) = (p(x))^t$ for some monic irreducible polynomial $p(x)$ and some integer $t \geq 1$. Further let $u \in V$ be such that the minimal polynomial of u relative to T is $f(x)$ itself then there exists a T -invariant subspace W of V such that $V = W \oplus F[T]u$.

Proof: Consider the family $\mathcal{F}$ of all those T -invariant subspaces U of V which have zero intersection with $F[T]u$. Since the subspace (0) is T -invariant and obviously $(0) \cap F[T]u = (0)$, $(0) \in \mathcal{F}$. Hence $\mathcal{F}$ is a non-empty family. Since no member of $\mathcal{F}$ can have dimension more than $n = \dim V$, we can find a member W of $\mathcal{F}$ having largest dimension among all members of $\mathcal{F}$. So W is a T -invariant subspace of V of largest dimension such that $W \cap F[T]u = (0)$.

We claim $V = W \oplus F[T]u$. Suppose the contrary. So there exists $v \in V$ such that $v \notin W + F[T]u$. Now $[F(T)]^0 v = Iv = v \notin W + F[T]u$, but $[p(T)]^t v = 0$ is in $W + F[T]u$. So we can find a positive integer $l \leq t$ such that $[p(T)]^l v \in W + F[T]u$, but $[p(T)]^{l-1} v \notin W + F[T]u$. Take $y = [p(T)]^{l-1} v$. Thus y is in V but not in $W + F[T]u$. Also $p(T)y = [p(T)]^l v$ is in $W + F[T]u$. So

$$p(T)y = w + g(T)u. \quad \dots(1)$$

for some $w \in W$ and $g(x) \in F[x]$. Further $0 = [p(T)]^t y = [p(T)]^{t-1} p(T)y$

$$= [p(T)]^{t-1} w + [p(T)]^{t-1} g(T)u. \quad \dots(2)$$

As the sum $W + F[T]u$ is direct we get $[p(T)]^{t-1} w = 0$ and $[p(T)]^{t-1} g(T)u = 0$. Hence as $[p(x)]^t$ is the minimal polynomial of u relative to T , we get $[p(x)]^t \mid [p(x)]^{t-1} g(x)$. This yields that $p(x) \mid g(x)$. So that $g(x) = p(x)g_1(x)$ for some $g_1(x) \in F[x]$. Then from (1) we have

$$p(T)y = w + p(T)g_1(T)u \quad \dots(3)$$

i.e., $p(T)[y - g_1(T)u] = w \in W$. Consider the T -invariant subspace $W_1 = W + F[T][y - g_1(T)u]$, $W \subseteq W_1$. If $y - g_1(T)u \in W$, then we shall get $y = g_1(T)u + w'$ for some $w' \in W$ and hence $y \in F(T)u + W$. This is a contradiction. Hence $y - g_1(T)u \notin W$. This prove that W_1 properly contains W . So $\dim W_1 > \dim W$.

Since, by choice W is a T -invariant subspace of largest dimension satisfying, $W \cap F[T]u = (0)$, we get $W_1 \cap F[T]u \neq (0)$. Thus for some non-zero $w_1 \in W_1 \subset F[T]u$

$$w_1 = w' + h(T)[y - g_1(T)u]; w' \in W \quad \dots(4)$$

As $w_1 \in F[T]u$, we get

$$w' + h(T)[y - g_1(T)u] = h'(T)u \quad \dots(5)$$

for some $h'(x) \in F[x]$. Therefore

$$h(T)y = -w' + [h(T)g_1(T) + h'(T)]u \in W + F[T]u \quad \dots(6)$$

If $p(x) \mid h(x)$, then $h(x) = p(x)h_1(x)$, $h_1(x) \in F[x]$. Consequently (3) and (4) yield $w_1 = w' + h_1(T)p(T)[y - g_1(T)u] \in W$. However by (4), $w_1 \in F[T]u$. So that $w_1 \in W \cap F[T]u = (0)$. This contradicts the fact that $w_1 \neq 0$. Hence $p(x)$ does not divide $h(x)$ i.e., the HCF of $p(x)$ and $h(x)$ is 1. Consequently $1 = a(x)h(x) + b(x)p(x)$ for some $a(x), b(x) \in F[x]$. Hence

$$I = a(T)h(T) + b(T)p(T).$$

Consequently $y = Iy = a(T)h(T)y + b(T)p(T)y$.

Since by (1), $p(T)y \in W + F[T]u$ and by (6), $h(T)y \in W + F[T]u$ and $W + F[T]u$ is a T -invariant subspace of V , the last equation yields that $y \in W + F[T]u$. This contradicts the supposition that $y \notin W + F[T]u$. Hence $V = W \oplus F[T]u$. ■

Theorem 16.52. Let $T \in A(V)$ be such that its minimal polynomial $f(x) = (p(x))^k$ for some irreducible monic polynomial $p(x)$ and some positive integer k . Then $V = F[T]u_1 \oplus F[T]u_2 \oplus \dots \oplus F[T]u_t$, for some non-zero vectors u_i ($1 \leq i \leq t$) such that the minimal polynomials of u_i 's relative to T are $(p(x))^{k_i}$'s satisfying the condition: $k = k_1 \geq k_2 \geq \dots \geq k_t \geq 1$. These k_i 's are uniquely determined by T .

Proof: V has a non-zero member u_1 such that the minimal polynomial of u_1 relative to T is $f(x) = (p(x))^{k_1}$; $k_1 = k$ (Theorem 16.31).

By Theorem 16.51, $V = F[T]u_1 \oplus V_1$ for some T -invariant subspace V_1 of V . If $V_1 = (0)$, we are done. Let $V_1 \neq (0)$. Clearly $\dim V_1 < \dim V$. Let T_1 be the restriction of T to V_1 , then the minimal polynomial of T_1 being a factor of $f(x) = (p(x))^k$ equals $(p(x))^{k_2}$ for some $k_2 \leq k_1$. If $k_2 = 0$, then $(p(x))^{k_2} = I$ yields that $0 = (p(T))^{k_2}V_1 = IV_1 = V_1$. This is a contradiction. Hence $k_2 \geq 1$. Again by Theorem 16.31, we can find a vector u_2 in V_1 such that the minimal polynomial of u_2 relative to T_1 (hence relative to T as T_1 is restriction of T) is $(p(x))^{k_2}$ and by Theorem 16.51, we get

$$V_1 = F[T]u_2 \oplus V_2, \quad \dots(2)$$

for some T_1 -invariant (hence T -invariant) subspace V_2 of V . Then $V = F[T]u_1 \oplus F[T]u_2 \oplus V_2$.

If V_2 is (0) we are done, otherwise considering the restriction T_2 of T_1 (hence of T) to V_2 we proceed as we did for V or V_1 . As

$\dim V_2 < \dim V_1 < \dim V$ and $\dim V$ is finite, this process ends after finitely many steps. So that we get

$$V = F[T]u_1 \oplus F[T]u_2 \oplus \dots \oplus F[T]u_t, \quad \dots(3)$$

such that the minimal polynomial of each of u_i is $(p(x))^{k_i}$ satisfying the condition $k = k_1 \geq k_2 \geq k_3 \geq \dots \geq k_t \geq 1$.

We want to show that $k_1, k_2, \dots, k_t$ are independent of the decomposition (3). For this, suppose that

$$V = F[T]u_1' \oplus F[T]u_2' \oplus \dots \oplus F[T]u_s', \quad \dots(4)$$

such that the minimal polynomial of each u_i' is $(p(x))^{k_i'}$ satisfying the condition $k = k_1' \geq k_2' \geq \dots \geq k_s' \geq 1$.

Consider $W = \{v \in V \mid p(T)v = 0\}$. W is a subspace of V ; further for any $w \in W$, $p(T)[T(w)] = T[p(T)w] = 0 \Rightarrow T(w) \in W$. Hence W is a T -invariant subspace of V . Let $w \in W$. By (3), $w = \sum_{i=1}^t g_i(T)u_i$ for

some $g_i(x) \in F[x]$, ($1 \leq i \leq t$). Hence $p(T)w = 0$ yield that

$$p(T)g_1(T)u_1 + p(T)g_2(T)u_2 + \dots + p(T)g_t(T)u_t = 0.$$

As (3) is a direct sum, we get $p(T)g_i(T)u_i = 0 \forall 1 \leq i \leq t$. This gives that $(p(x))^{k_i} \mid p(x)g_i(x)$ as $(p(x))^{k_i}$ is the minimal polynomial of u_i .

Consequently $(p(x))^{k_i-1} \mid g_i(x)$ i.e., $g_i(x) = h_i(x)(p(x))^{k_i-1}$. So if we put $w_i = (p_i(T))^{k_i-1}u_i$, we see that $0 \neq w_i \in W$ and every member of W is of the form

$$h_1(T)w_1 + h_2(T)w_2 + \dots + h_t(T)w_t.$$

Hence $W = F[T]w_1 + F[T]w_2 + \dots + F[T]w_t$. Further as $F[T]w_i \subseteq F[T]u_i$, ($1 \leq i \leq t$) and by (3) the sum of the $F[T]u_i$'s is direct, the sum of $F[T]w_i$'s is also direct. In other words

$$W = F[T]w_1 \oplus F[T]w_2 \oplus \dots \oplus F[T]w_t. \quad \dots(5)$$

Similarly using (4) we get

$$W = F[T]w_1' \oplus F[T]w_2' \oplus \dots \oplus F[T]w_s', \quad \dots(6)$$

where

$$0 \neq w_i' = [p(T)]^{k_i'-1}u_i', \quad 1 \leq i \leq s.$$

Since any non-zero $w \in W$ satisfies $p(T)w = 0$, the minimal polynomial of w is a factor of $p(x)$. Again since the minimal polynomial of a non-zero vector is always of positive degree, and $p(x)$ is irreducible, we get $p(x)$ is itself is the minimal of w . Hence the minimal polynomial of every w_i or w_j' is $p(x)$. Consider $K = \frac{F[x]}{\langle p(x) \rangle}$.

K is a field (why?). For any coset $\overline{g(x)} = g(x) + \langle p(x) \rangle \in K$ and any $w \in W$, define $\overline{g(x)}w = g(T)w$.

This is well-defined. For let $\overline{g(x)} = \overline{g_1(x)}$ for some $g(x), g_1(x) \in F[x]$. Then $g(x) - g_1(x) \in \langle p(x) \rangle$ i.e., $p(x) \mid [g(x) - g_1(x)]$. Then as $p(T)w = 0$ we get $[g(T) - g_1(T)]w = 0$ i.e., $g(T)w = g_1(T)w$ i.e., $\overline{g(x)}w = \overline{g_1(x)}w$. It is now

This shows that the product $\overline{g(x)}w$ is well defined.

straight forward to verify that this definition makes W into a vector space over K . Clearly $Kw = F[x]w = F[T]w$. Hence by (5), $W = Kw_1 \oplus Kw_2 \oplus \dots \oplus Kw_t$.

This shows that $\dim_K W = t$. Similarly by (6) $\dim_K W = s$. Hence $t = s$. This proves that the number t is uniquely determined by T .

So it remains to show that $k_1 = k_1', k_2 = k_2', \dots, k_t = k_t'$. For this we apply induction on the dimension of V . So we suppose that the result holds for all vector spaces of dimension $< \dim V$. Consider $V' = p(T)V$. Since $T(V') = p(T)[TV] \subseteq p(T)V$, V' is a T -invariant subspace of V . Further $[p(T)]^{k_1-1}V' = p(T)^{k_1}V = (0)$, but $[p(T)]^{k_1-1}V' \neq (0)$, we get $V' \neq V$, i.e., $\dim V' < \dim V$. So if T' is the restriction of T to V' , then, by induction hypothesis, the result holds for V' and T' . Since $k = k_1 \geq k_2 \geq \dots \geq k_t \geq 1$, we can find integer $\alpha \geq 0$ such that $k_i = 1$ for all $i > \alpha$ and $k_i > 1$ for $i \leq \alpha$. Then for $i > \alpha$, $p(T)u_i = [p(T)]^{k_i}u_i = 0$, but for $i \leq \alpha$, $p(T)u_i \neq 0$. Again by (3), $V' = p(T)V = F[T]p(T)u_1 \oplus \dots \oplus F[T]p(T)u_\alpha$

$$= F[T]v_1 \oplus \dots \oplus F[T]v_\alpha \quad \dots(7)$$

where $v_i = p(T)u_i$ ($1 \leq i \leq \alpha$), has $[p(x)]^{k_i-1}$ as its minimal polynomial relative to T (hence relative to T'). Similarly we get

$$V' = F[T]v_1' \oplus F[T]v_2' \oplus \dots \oplus F[T]v_\beta \quad \dots(8)$$

where β is an integer such that $k_i' = 1$ for $i > \beta$ and $k_i' > 1$ for $i \leq \beta$, and $v_i' = p(T)u_i'$ has $[p(x)]^{k_i'-1}$ as its minimal polynomial relative to T (hence relative to T').

By induction hypothesis we have $\alpha = \beta$, $k_i - 1 = k_i' - 1$ for $i \leq \alpha$, i.e., $k_i = k_i'$ for $i \leq \alpha$. Obviously as $k_i = 1$ and $k_i' = 1$ for $i > \alpha$, $k_i = k_i'$ for $i > \alpha$. This proves that $k_i = k_i'$ for all i . Thus the theorem is proved. ■

Theorem 16.53. Let T be a linear transformation on V having $f(x) = [p_1(x)]^{k_1} [p_2(x)]^{k_2} \dots [p_r(x)]^{k_r}$ as its minimal polynomial, where $p_1(x), p_2(x), \dots, p_r(x)$ are distinct monic irreducible polynomials and all k_i 's are positive integers. Then there exist unique T -invariant subspaces $V_1, V_2, \dots, V_r$ of V such that $V = V_1 \oplus V_2 \oplus \dots \oplus V_r$, $T = T_1 \oplus T_2 \oplus \dots \oplus T_r$ where T_i is the restriction of T to V_i and has $[p_i(x)]^{k_i}$ as its minimal polynomial.

Proof. Let $g_i(x) = \frac{f(x)}{[p_i(x)]^{k_i}}$ for $1 \leq i \leq r$, $g_i(x)$ is the product of all those $[p_j(x)]^{k_j}$ for which $j \neq i$;

Define $V_i = g_i(T)V$; $1 \leq i \leq r$... (1)

V_i is a T -invariant subspace of V . Since $[p_i(T)]^{k_i} V_i = [p_i(T)]^{k_i} g_i(T)V = f(T)V = (0)$, we get that the minimal polynomial of the restriction T_i of T is $[p_i(x)]^{k_i}$ [?]. Since the HCF of $g_1(x), g_2(x), \dots, g_r(x)$ is 1, there exist polynomials $q_1(x), q_2(x), \dots, q_r(x)$ such that

$$I = g_1(x)q_1(x) + g_2(x)q_2(x) + \dots + g_r(x)q_r(x), \quad \dots(2)$$

$$\text{Hence } I = g_1(T)q_1(T) + g_2(T)q_2(T) + \dots + g_r(T)q_r(T) \quad \dots(3)$$

Then for any $v \in V$

$$\begin{aligned} v = I(v) &= g_1(T)q_1(T)v + g_2(T)q_2(T)v + \dots + g_r(T)q_r(T)v \\ &= v_1 + v_2 + \dots + v_r, \text{ where } v_i = g_i(T)q_i(T)v \in V_i = g_i(T)V. \end{aligned}$$

Hence $V = V_1 + V_2 + \dots + V_r$.

To show that the sum is direct, let for some $x_i \in V_i$ ($1 \leq i \leq r$)

$$x_1 + x_2 + \dots + x_r = 0 \quad \dots(4)$$

Since for any i , $[p_j(x)]^{k_j} \mid g_i(x)$ for $j \neq i$ and $[p_j(T)]^{k_j} V_j = (0)$, we have $g_i(T)V_j = (0)$ for $i \neq j$. So from (4) we get

$$g_i(T)x_i = 0 \quad \dots(5)$$

Now $g_i(x)$, $[p_i(x)]^{k_i}$ are relatively prime, so for polynomials $a(x)$ and $b(x)$ in $F[x]$,

$$1 = a(x)g_i(x) + b(x)[p_i(x)]^{k_i} \quad \dots(6)$$

$$I = a(T)g_i(T) + b(T)[p_i(T)]^{k_i} \quad \dots(7)$$

So that

$$I = a(T)g_i(T) + b(T)[p_i(T)]^{k_i}$$

Consequently $x_i = I(x_i) = a(T)g_i(T)x_i + b(T)[p_i(x)]^{k_i}x_i = 0$, since $g_i(T)x_i = 0$ [by (5)] and $[p_i(T)]^{k_i}V_i = (0)$. Hence in (4) every $x_i = 0$.

So that

$$V = V_1 \oplus V_2 \oplus \dots \oplus V_r \quad \dots(8)$$

To show the uniqueness of these V_i 's let also

$$V = W_1 \oplus W_2 \oplus \dots \oplus W_r \quad \dots(9)$$

for some T -invariant subspaces W_i ($1 \leq i \leq r$) such that the minimal polynomial of the restriction T_i of T to W_i is $[p_i(x)]^{k_i}$. Then as $[p_i(x)]^{k_i} \mid g_i(x)$ for every $i \neq j$, we have $g_i(T)W_j = (0)$ for $i \neq j$, so from (9) we get

$$g_i(T)V = g_i(T)W_i \subseteq W_i; \text{ i.e. } V_i \subseteq W_i.$$

However given any $w_i \in W_i$, because of (7), we have

$$\begin{aligned} w_i &= g_i(T)a(T)w_i + b(T)[p_i(T)]^{k_i}w_i \\ &= g_i(T)a(T)w_i \in g_i(T)V = V_i. \end{aligned}$$

Hence $W_i \subseteq V_i$. Consequently $W_i = V_i$.

This proves that V_i 's are uniquely determined. ■

REMARK: We claim that $V_i = \{v \in V \mid [p_i(T)]^{\alpha_i} v = 0 \text{ for some } \alpha_i \geq 1\}$. Since $[p_i(T)]^{k_i} V_i = (0)$, $[p_i(T)]^{k_i} v = 0$ for all $v \in V_i$. Conversely let $v \in V$ satisfy $[p_i(T)]^{\alpha_i} v = 0$ for some $\alpha_i \geq 1$. Since the HCF of $g_i(x)$ and $[p_i(x)]^{\alpha_i}$ is 1, $1 = g_i(x)a_1(x) + [p_i(x)]^{\alpha_i}b_1(x)$ for some $a_1(x), b_1(x) \in F[x]$. Hence $v = Iv = g_i(T)a_1(T)v + b_1(T)[p_i(T)]^{\alpha_i}v = g_i(T)a_1(T)v \in V_i$. Hence $V_i = \{v \in V \mid [p_i(T)]^{\alpha_i} v = 0 \text{ for some } \alpha_i \geq 1\}$.

Theorem 16.54. (Fundamental Theorem on Direct Sums). Let V be any vector space of dimension $n > 0$ over a field F and T , a linear transformation on V . Then there exist unique polynomials $q_1(x), q_2(x), \dots, q_m(x)$, each a power of monic polynomial, and cyclic T -invariant sub-

spaces $W_1, W_2, \dots, W_m$ (not necessarily unique) such that $V = W_1 \oplus W_2 \oplus \dots \oplus W_m$ and the restriction T_i of T to W_i has $q_i(x)$ as its minimal polynomial.

Proof: Let $f(x) = [p_1(x)]^{k_1} [p_2(x)]^{k_2} \dots [p_r(x)]^{k_r}$ be the minimal polynomial of T where $p_1(x), p_2(x), \dots, p_r(x)$ are distinct monic irreducible polynomials over F and k_i are positive integers. In view of Theorem 16.53 and Remark following it, we get

$$V = V_1 \oplus V_2 \oplus \dots \oplus V_r, \quad \dots (1)$$

where each V_i is a T -invariant subspace of V such that the restriction L_i of T to V_i has $[p_i(x)]^{k_i}$ as its minimal polynomial, further $V_i = \{v \in V \mid [p_i(x)]^{\alpha_i} v = 0 \text{ for some } \alpha_i \geq 1\}$. Now by Theorem 16.52 we can write

$$V_i = F[T]u_{i1} \oplus F[T]u_{i2} \oplus \dots \oplus F[T]u_{it_i} \quad \dots (2)$$

for some non-zero vectors u_{ij} ($1 \leq j \leq t_i$) in V_i such that the minimal polynomial of u_{ij} relative to L_i is $[p_i(x)]^{k_{ij}}$ satisfying the condition, $k_i = k_{i1} \geq k_{i2} \geq \dots \geq k_{it_i}$. These k_{ij} 's are determined uniquely by L_i . As L_i is the each restriction of T to V_i , we also get that the minimal polynomial of each u_{ij} relative to T is $[p_i(x)]^{k_{ij}}$. Hence if we write

$$W_{ij} = F[T]u_{ij}, \quad 1 \leq j \leq t_i, \quad 1 \leq i \leq r, \quad \dots (3)$$

then from (1) and (2) we get

$$V = W_{11} \oplus W_{12} \oplus \dots \oplus W_{1t_1} \oplus \dots \oplus W_{r1} \oplus W_{r2} \oplus \dots \oplus W_{rt_r} \quad \dots (4)$$

Each W_{ij} is cyclic T -invariant subspace such that the restriction T_{ij} of T to W_{ij} has $[p_i(x)]^{k_{ij}}$ as its minimal polynomial, which is a power of an irreducible monic polynomial. Hence the theorem will follow if we show that these monic polynomials are uniquely determined by T . To this effect let

$$V = W_1 \oplus W_2 \oplus \dots \oplus W_m \quad \dots (5)$$

where each W_i is a cyclic T -invariant subspace such that the restriction of T to W_i has $q_i(x)$ as its minimal polynomial and $q_i(x)$ is a power of some irreducible monic polynomial. Then $f(x) = [p_1(x)]^{k_1} \dots [p_r(x)]^{k_r}$ is the LCM of $q_1(x), q_2(x), \dots, q_m(x)$. Hence each $q_i(x) = [p_{j(i)}(x)]^{\alpha_i}$ for some α_i , $1 \leq \alpha_i \leq k_{j(i)}$ and for some i , $\alpha_i = k_{j(i)}$. By Remark following Theorem 16.53 we get $W_i \subseteq V_{j(i)}$. Further for each $i = 1, 2, \dots, r$, some $q_j(x)$ must be a power of $p_i(x)$. Hence if for each $i = 1, 2, \dots, r$ we take U_i as the sum of those W_j 's for which $q_j(x)$ is a power of $p_i(x)$, we get $U_i \subseteq V_i$, therefore

$$V = U_1 \oplus U_2 \oplus \dots \oplus U_r \quad \dots (6)$$

with each U_i being a T -invariant subspace of V and $U_i \subseteq V_i$. Consider any $x \in V_i$, by (6), $x = u_1 + u_2 + \dots + u_r$. This yields $x - u_i = \sum_{j \neq i} u_j \in V_i \cap (\sum_{j \neq i} V_j) = (0)$ since $u_j \in U_j \subseteq V_j$ for all j . Hence $x = u_i \in U_i$. As a

consequence $U_i = V_i$ for $1 \leq i \leq r$.

Now U_i (hence V_i) is a direct sum of those cyclic space W_j 's for which the restriction of T to W_i has $q_j(x)$ a power of $p_i(x)$; because of Theorem 16.52, the number of these $w_j \in W_i$ is equal to t_i [from (2)] and these $q_j(x)$'s are equal to $[p_i(x)]^{k_{ij}}$ ($1 \leq j \leq t_i$). Hence $q_j(x)$'s are determined uniquely by T . This proves the theorem. ■

Definition 16.55. Let T be a linear transformation on a vector space V and $V = W_1 \oplus W_2 \oplus \dots \oplus W_m$ be a decomposition of V into cyclic T -invariant subspaces, such that the restriction T_i of T to each W_i has a minimal polynomial $q_i(x)$ which is a power of some monic irreducible polynomial.

$q_1(x), q_2(x), \dots, q_m(x)$ are called *elementary divisors* of T . If A is a matrix of T relative to some ordered basis of V , then the elementary divisors of T are also called the *elementary divisors* of A .

REMARK. These $q_i(x)$'s need not be distinct. We shall see this below, though by Theorem 16.54, $q_i(x)$ are uniquely determined by T .

Corollary 16.56. Any root in F of an elementary divisor of a linear transformation T on V_F is an eigen value of T .

Proof: Now $V = W_1 \oplus W_2 \oplus \dots \oplus W_m$, in the notation of above definition. Since W_i is a cyclic T_i -space, there exists $w \in W_i$ whose minimal polynomial relative to T_i (hence relative to T) is $q_i(x)$. Consider any root α of $q_i(x)$ in F . Then $q_i(x) = g(x)(x - \alpha)$ for some $g(x) \in F[x]$. As $\deg g(x) < \deg q_i(x)$, $[g(T)]w \neq 0$. Put $v = g(T)w$. Then $(T - \alpha)v = (T - \alpha)[g(T)w] = q_i(T)w = 0$ yields $T(v) = \alpha v$. Hence α is an eigen value of T . ■

EXAMPLE 19. Consider $\mathbb{Q}^4$. Let T be a linear transformation on $\mathbb{Q}^4$ such that $T(e_1) = e_2$, $T(e_2) = 0$, $T(e_3) = e_4$, $T(e_4) = 0$ where (e_1, e_2, e_3, e_4) is the standard basis of $\mathbb{Q}^4$. Clearly $T^2 = 0$, but $T \neq 0$. So that the minimal polynomial of T is x^2 , which is a power of $p(x) = x$; $p(x)$ is trivially irreducible.

Now $\mathbb{Q}^4 = \mathbb{Q}[T]e_1 \oplus \mathbb{Q}[T]e_3$, a direct sum of two cyclic subspaces. Since the minimal polynomial of each of e_1, e_3 relative to T is x^2 , we can say that the restrictions T_1, T_3 of T to $\mathbb{Q}[T]e_1$ and $\mathbb{Q}[T]e_3$ have x^2 as their respective minimal polynomials. Hence in the terminology of Definition 16.55, x^2, x^2 are the two elementary divisors of T , they happen to be equal is this case.

EXAMPLE 20. Consider the matrix

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 2 \end{pmatrix}$$

Let T be the linear transformation on $\mathbb{Q}^3$ such that A is the

matrix of T relative to an ordered basis (v_1, v_2, v_3) of $\mathbf{Q}^3$. Then $T(v_1)=v_1$, $T(v_2)=v_3$ and $T(v_3)=-v_2+2v_3$. Clearly the minimal polynomial of v_1 relative to T is $x-1$; which being of degree 1 yields that $F[T]v_1=\mathbf{Q}v_1$.

$(T-I)v_2=(T^2-2T+I)v_2=0$, but $(T-I)v_2\neq 0$. Thus as $x-1$ is irreducible, we get $(x-1)^2$ is the minimal polynomial of v_2 relative to T . Hence $\mathbf{Q}(T)v_2$ is the space generated by $[v_2, T(v_2)]$ i.e. by (v_2, v_3) . Hence $\mathbf{Q}[T]v_2=\mathbf{Q}v_2+\mathbf{Q}v_3$. Consequently $\mathbf{Q}^3=\mathbf{Q}v_1\oplus(\mathbf{Q}v_2\oplus\mathbf{Q}v_3)=\mathbf{Q}[T]v_1\oplus\mathbf{Q}[T]v_2$.

Thus $x-1$ $(x-1)^2$ are the elementary divisors of T .

EXAMPLE 21. Consider the 5×5 matrix.

$$A=\begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 & 4 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

Let T be a linear transformation on $\mathbf{Q}^5$ whose matrix relative to the standard basis $(e_1, e_2, e_3, e_4, e_5)$ of $\mathbf{Q}^5$, is A .

$$\text{If } A_1=\begin{pmatrix} 0 & 1 \\ 1 & 2 \end{pmatrix}, \quad A_2=\begin{pmatrix} 0 & 0 & -1 \\ 1 & 0 & 4 \\ 0 & 1 & 0 \end{pmatrix}$$

then as $A=A_1\oplus A_2$, we get that $W_1=\langle e_1, e_2 \rangle$, $W_2=\langle e_3, e_4, e_5 \rangle$ are T -invariant subspaces of V and $V=W_1\oplus W_2$.

Let T_i be the restrictions of T to W_i ($i=1, 2$). Then the matrix of T_i is A_i . Now A_1 is the companion matrix of the irreducible polynomial x^2-2x-1 . Since $T(e_1)=e_2$ we get that $W_1=F[T]e_1$ is a cyclic subspace and T_1 has x^2-2x-1 as its minimal polynomial. Again as $T(e_3)=e_4$, $T(e_4)=e_5$, we also have $W_2=F[T]e_3$ a cyclic subspace. A_2 being the companion matrix of the irreducible polynomial x^3-4x+1 yields that the minimal polynomial of T_2 is x^3-4x+1 . Hence the minimal polynomial of T is $(x^2-2x-1)(x^3-4x+1)$, and the elementary divisors of T are x^2-2x-1 and x^3-4x+1 .

REMARK. Given finitely many monic polynomials $q_1(x), q_2(x), \dots, q_t(x)$ over a field F ; each of which is a power of some irreducible polynomial over F , we can construct a linear transformation T on a vector space V_F of dimension equal to the sum of the degrees of the given polynomials, such that $q_1(x), q_2(x), \dots, q_t(x)$ are elementary divisors of T . For this let A_i ($1\leq i\leq t$) be the companion matrix of $q_i(x)$. Let $m_i=\deg q_i(x)$ and W_i be a vector space over F of dimension m_i . Then there exists a linear transformation T_i on W_i whose matrix relative to an ordered basis $\mathbf{B}_i$ is A_i and its minimal

polynomial is $q_i(x)$. (Theorems 16.35 and 16.36). As seen in the proof of Theorem 16.35, W_i is a cyclic T_i -space. Then if $V = W_1 \oplus W_2 \oplus \dots \oplus W_t$ and $T = T_1 \oplus T_2 \oplus \dots \oplus T_t$, we have that each W_i is a cyclic T -subspace of V such that T_i equals the restriction of T to W_i and $q_i(x)$ is the minimal polynomial of T_i . Hence $q_1(x), q_2(x), \dots, q_t(x)$ are the elementary divisors of T .

EXAMPLE 22. Let $(x+1), (x+1)^2, (x+1), (x^2+x+1)$ be four polynomials each of which is clearly a power of some irreducible polynomial over $\mathbf{Q}$. Their respective companion matrices are

$$A_1 = (-1), \quad A_2 = \begin{pmatrix} 0 & -1 \\ 1 & -2 \end{pmatrix}, \quad A_3 = (-1), \quad A_4 = \begin{pmatrix} 0 & -1 \\ 1 & -1 \end{pmatrix}$$

$$\text{Now } A_1 \oplus A_2 \oplus A_3 \oplus A_4 = \begin{pmatrix} -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 1 & -2 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 & 1 & -1 \end{pmatrix}$$

So if any linear transformation T on $\mathbf{Q}^6$ has its matrix similar to the above matrix, then $(x+1), (x+1)^2, (x+1), (x^2+x+1)$ are its elementary divisors.

PROBLEMS

1. In each of the following exercises, show that the minimal polynomial of T is a power of an irreducible polynomial, and then represent the vector space V as a direct sum of T -cyclic subspaces.

- $V = \mathbf{Q}^2$, $T(e_1) = 2e_1$, $T(e_2) = 2e_2$.
- $V = \mathbf{Q}^2$, $T(e_1) = -e_2$, $T(e_2) = e_1$.
- $V = \mathbf{Q}^2$, $T(e_1) = e_1 - e_2$, $T(e_2) = 2e_1 - e_2$.
- $V = \mathbf{Q}^3$, $T(e_1) = e_1$, $T(e_2) = 2e_1 - e_2 + e_3$, $T(e_3) = 4e_1 - 4e_2 + 3e_3$.
- $V = \mathbf{Q}^3$, $T(e_1) = e_2$, $T(e_2) = e_3$, $T(e_3) = -e_1 + 4e_2$.
- $V = \mathbf{Q}^3$, $T(e_1) = e_2$, $T(e_2) = e_3$, $T(e_3) = 2e_1$.
- $V = \mathbf{Q}^4$, $T(e_1) = -e_2$, $T(e_2) = e_1$, $T(e_3) = -e_4$, $T(e_4) = e_3$.
- $V = \mathbf{Q}^4$, $T(e_1) = e_1$, $T(e_2) = e_2$, $T(e_3) = 2e_2 - e_3 + e_4$, $T(e_4) = 4e_2 - 4e_3 + 3e_4$.
- $V = \mathbf{Q}^4$, $T(e_1) = -e_2 + 2e_3 - 2e_4$, $T(e_2) = -3e_1 + 4e_2 - 2e_3 - 6e_4$, $T(e_3) = -2e_2 + 2e_3 + e_4$, $T(e_4) = -2e_1 + 4e_2 - 3e_3 - 6e_4$.

2. Prove that for any linear transformation T on a vector space V_F , the sum of the degrees of the elementary divisors of T equals the dimension of V_F .

3. Prove that any two linear transformations T and T' on a vector space V , have same elementary divisors, if and only if their matrices relative to a basis of V are similar.

4. Prove that the minimal polynomial of a linear transformation T is a factor of the product of its elementary divisors and each elementary divisor is a factor of the minimal polynomial.

5. Find the elementary divisors of each of the following linear transformations T , and express the vector space V as a direct sum of cyclic T -invariant subspaces.

(a) $V = \mathbf{Q}^2$, $T(e_1) = e_1$, $T(e_2) = 3e_2$

(b) $V = \mathbf{Q}^2$, $T(e_1) = e_2$, $T(e_2) = e_1$

(c) $V = \mathbf{Q}^2$, $T(e_1) = 5e_2$, $T(e_2) = 3e_2$

(d) $V = \mathbf{Q}^3$, $T(e_1) = e_1 + e_2 + e_3$, $T(e_2) = -e_3$, $T(e_3) = 2e_2 + 3e_3$

(e) $V = \mathbf{Q}^3$, $T(e_1) = 2e_1 - e_2 - 2e_3$, $T(e_2) = -6e_1 - 3e_2 + 2e_3$,
 $T(e_3) = 8e_1 + 3e_2 - 3e_3$.

(f) $V = \mathbf{Q}^4$, $T(e_1) = -e_2$, $T(e_2) = e_1$, $T(e_3) = e_4$, $T(e_4) = e_3$.

(g) $V = \mathbf{Q}^4$, $T(e_1) = e_2$, $T(e_2) = e_1 + e_2 + e_3$, $T(e_3) = e_4$, $T(e_4) = e_3 + e_4$.

6. If a linear transformation T on a vector space V_F has $(p(x))^k$ as its minimal polynomial, where $p(x)$ is irreducible, then what possible distinct elementary divisors T can have?

5. Jordan Normal Form

Let $A = (a_{ij})$ be an $n \times n$ matrix over a field F and V be an n -dimensional vector space over F with $\mathbf{B} = (b_1, b_2, \dots, b_n)$ as its ordered basis. Then the linear transformation T on V defined by $T(b_j) = \sum_{i=1}^n a_{ij} b_i$, $1 \leq j \leq n$ admits A as its matrix relative to $\mathbf{B}$. Let $q_1(x)$,

$q_2(x), \dots, q_m(x)$ be the elementary divisors of T . Then $V = W_1 \oplus W_2 \oplus \dots \oplus W_m$, $T = T_1 \oplus T_2 \oplus \dots \oplus T_m$ for some cyclic T -invariant subspaces W_i ($1 \leq i \leq m$) such that the restriction T_i of T to W_i has $q_i(x)$ as its minimal polynomial (Theorem 16.54). Since each W_i is cyclic T_i -space by Theorem 16.26, W_i has a basis $\mathbf{B}_i$, such that the matrix of T_i relative to $\mathbf{B}_i$ is the Jordan matrix of $q_i(x)$ relative to the irreducible polynomial of which $q_i(x)$ is a power.

$T = T_1 \oplus T_2 \oplus \dots \oplus T_m$ yields that V has a basis $\mathbf{B}' = (b'_1, b'_2, \dots, b'_n)$ such that the matrix of T relative to $\mathbf{B}'$ is $J = J_1 \oplus J_2 \oplus \dots \oplus J_m$. Then A is similar to J (Corollary 12.40). We call J , a *Jordan Normal Form* of matrix A . Since the elementary divisors of T are uniquely determined (Theorem 16.54), the Jordan Normal Form of A is uniquely determined except for the orders in which J_i 's are written.

If a polynomial $q(x)$ is a power of a monic irreducible polynomial $p(x)$, then the Jordan matrix of $q(x)$ relative to $p(x)$ will be simply called the *Jordan matrix of $q(x)$* .

Theorem 16.57. For any $A \in M_n(\mathbf{C})$, Jordan Normal Form of the matrix A is of the type

$$\begin{pmatrix} J_1 & O & O & \dots & O \\ O & J_2 & O & \dots & O \\ O & O & J_3 & \dots & O \\ \dots & & & \ddots & \\ O & O & O & \dots & J_m \end{pmatrix}$$

where each J_i is of the form

$$\begin{pmatrix} \alpha_i & 1 & 0 & \dots & 0 & 0 \\ 0 & \alpha_i & 1 & \dots & 0 & 0 \\ 0 & 0 & \alpha_i & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & \alpha_i & 1 \\ 0 & 0 & 0 & \dots & 0 & \alpha_i \end{pmatrix}$$

Further each α_i is an eigen value of A .

Proof: Let $q_1(x), \dots, q_m(x)$ be the elementary divisors of A . Since over $\mathbb{C}$ only irreducible polynomials are of degree one, and each $q_i(x)$ is a power of an irreducible polynomial, we have $q_i(x) = (x - \alpha_i)^{k_i}$ for some $\alpha_i \in \mathbb{C}$ and some positive integer k_i . Then the Jordan matrix of $q_i(x)$ is the $k_i \times k_i$ matrix

$$J_i = \begin{pmatrix} \alpha_i & 1 & 0 & \dots & 0 & 0 \\ 0 & \alpha_i & 1 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & \alpha_i & 1 \\ 0 & 0 & 0 & \dots & 0 & \alpha_i \end{pmatrix}$$

Hence the Jordan normal form of A is

$$J_1 \oplus J_2 \oplus \dots \oplus J_m = \begin{pmatrix} J_1 & O & \dots & O \\ O & J_2 & \dots & O \\ \dots & \dots & \dots & \dots \\ O & O & \dots & J_m \end{pmatrix}$$

The next part follows from Corollary 16.56. ■

EXAMPLE 23. Consider polynomials $(x+1)$, $(x+1)^2$, $(x-i)^2$, $(x+i)^2$ over $\mathbb{C}$. Their respective Jordan matrices (Jordan Blocks) are

$$J_1 = (-1), \quad J_2 = \begin{pmatrix} -1 & 1 \\ 0 & -1 \end{pmatrix}, \quad J_3 = \begin{pmatrix} i & 1 \\ 0 & i \end{pmatrix} \quad \text{and} \quad J_4 = \begin{pmatrix} -i & 1 \\ 0 & -i \end{pmatrix}.$$

The 7×7 matrix

$$\begin{pmatrix} J_1 & O & O & O \\ O & J_2 & O & O \\ O & O & J_3 & O \\ O & O & O & J_4 \end{pmatrix} = \begin{pmatrix} -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & i & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & i & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -i & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & -i \end{pmatrix}$$

is in Jordan Normal Form and its elementary divisors are the above

given polynomials. Further 1, -1 , i , $-i$ are its distinct eigen values.

Definition 16.58. A linear transformation T on a vector space V is said to be *nilpotent* of index $k \geq 1$ if $T^k = 0$ but $T^{k-1} \neq 0$.

Analogously, definition of a nilpotent matrix can be given. It is immediate that a linear transformation T is nilpotent of index k if and only if its corresponding matrix is nilpotent of index k . For example any $n \times n$ matrix, whose elements on and below the diagonal are zeros is nilpotent of index $\leq n$.

Theorem 16.59. If $T \in A(V)$ is nilpotent of index k , then there exists an ordered basis $\mathbf{B}$ of V such that the matrix of T relative to $\mathbf{B}$ is of the form

$$\begin{pmatrix} M_{k_1} & O & O & O \\ O & M_{k_2} & O & O \\ \dots & \dots & \dots & \dots \\ O & O & O & M_{k_r} \end{pmatrix}$$

where $k = k_1 \geq k_2 \geq \dots \geq k_r$ and for any positive integer t , M_t is $t \times t$ matrix given by

$$M_t = \begin{pmatrix} 0 & 0 & 0 & \dots & 0 & 0 \\ 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 1 & 0 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 1 & 0 \end{pmatrix}$$

Proof: Since $T^k = 0$, the minimal polynomial $m(x)$ of T divides x^k , hence $m(x) = x^l$ for some $l \leq k$. As $T^{k-1} \neq 0$, $l \leq k-1$. Hence $l = k$. So that $m(x) = x^k$. Let $q_1(x), q_2(x), \dots, q_r(x)$ be the elementary divisors of T . Since each of them divides the minimal polynomial $m(x) = x^k$, $q_i(x) = x^{k_i}$ for some $k_i \leq k$. So we can re-arrange $q_i(x)$'s in such a way that $k_1 \geq k_2 \geq \dots \geq k_r$. As the LCM of these $q_i(x)$'s is x^k , we get $k = k_1$. Now the Jordan matrix of $q_i(x) = x^{k_i}$ relative to x is by Example 15, the $k_i \times k_i$ matrix

$$M_{k_i} = \begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 1 \\ 0 & 0 & 0 & \dots & 0 \end{pmatrix}$$

Since any matrix is similar to its Jordan Normal matrix, there has an ordered basis $\mathbf{B}$ such that the matrix of T relative to $\mathbf{B}$ is

$$\begin{pmatrix} M_{k_1} & O & O \\ O & M_{k_2} & O \\ O & O & M_{k_r} \end{pmatrix} \blacksquare$$

REMARK. Since the elementary divisors of T above, are uniquely

determined, the integers $k_1, k_2, \dots, k_r$ in the above theorem are uniquely determined by T . These integers are called the *invariants* of the nilpotent linear transformation T .

EXAMPLE 23. Consider the $n \times n$ matrix

$$M = \begin{pmatrix} 0 & 0 & \dots & 0 & 0 \\ 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 & 0 \end{pmatrix}.$$

This matrix is the companion matrix of $f(x) = x^n$. The minimal polynomial of M is x^n , which being of degree n gives that M has only one elementary divisor namely x^n itself. Now the Jordan matrix of x^n relative to x is the $n \times n$ matrix

$$J = \begin{pmatrix} 0 & 1 & 0 & 0 & \dots & 0 \\ 0 & 0 & 1 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 & \dots & 1 \\ 0 & 0 & 0 & 0 & \dots & 0 \end{pmatrix}.$$

Hence the Jordan Normal Form of M is J . Its only invariant is n .

EXAMPLE 24. Consider the polynomial x^3 . Let us determine all Jordan Normal Forms for 6×6 matrices having x^3 as their minimal polynomial. All such matrices are nilpotent of index 3. Let A be any such matrix. If its invariants are $k_1, k_2, \dots, k_r$, then we have $3 = k_1 \geq k_2 \geq \dots \geq k_r$ and $k_1 + k_2 + \dots + k_r = 6$. 6 can be expressed in this form only in the following three ways

$$3+3, \quad 3+2+1, \quad 3+1+1+1.$$

Hence the only possible Jordan normal forms are

$$A_1 = \begin{pmatrix} 0 & 1 & 0 & \vdots & 0 & 0 & 0 \\ 0 & 0 & 1 & \vdots & 0 & 0 & 0 \\ 0 & 0 & 0 & \vdots & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & \vdots & 0 & 1 & 0 \\ 0 & 0 & 0 & \vdots & 0 & 0 & 1 \\ 0 & 0 & 0 & \vdots & 0 & 0 & 0 \end{pmatrix}$$

$$A_2 = \begin{pmatrix} 0 & 1 & 0 & \vdots & 0 & 0 & 0 \\ 0 & 0 & 1 & \vdots & 0 & 0 & 0 \\ 0 & 0 & 0 & \vdots & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & \vdots & 0 & 1 & 0 \\ 0 & 0 & 0 & \vdots & 0 & 0 & 0 \\ 0 & 0 & 0 & \vdots & 0 & 0 & 0 \end{pmatrix}$$

$$A_3 = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 0 & 1 & \cdots & 0 & 0 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 \end{pmatrix}.$$

We advise the readers to determine 8×8 Jordan Normal matrices which are nilpotent of indices 3, 4, 5, 6 and 7.

EXAMPLE 25. Consider the matrix

$$A = \begin{pmatrix} 1 & 1 & 1 \\ -1 & -1 & -1 \\ 1 & 1 & 0 \end{pmatrix}.$$

$A^3 = (0)$, but $A^2 \neq (0)$. A is nilpotent of index 3. Since only 3×3 matrix in Jordan Normal Form, which is nilpotent of index 3, is

$$J = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}. \quad A \text{ is similar to } J.$$

Consider the matrix $B = \begin{pmatrix} 1 & 1 & 1 \\ -1 & -1 & -1 \\ 0 & 0 & 0 \end{pmatrix}$. It is nilpotent of index 2. Clearly A and B can not be similar. Since the only way to write $3 = k_1 + k_2 + \dots + k_r$, where k_i are positive integers satisfying $2 = k_1 \geq k_2 \geq \dots \geq k_r$, is $3 = 2 + 1$, we get that the Jordan Normal form

$$\text{of } B \text{ is } \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

PROBLEMS

1. Show that two linear transformations in $A(V)$ are similar if and only if they have same families of elementary divisors.

2. Let $T \in A(V)$ have $f(x) = [p_1(x)]^{m_1} [p_2(x)]^{m_2} \dots [p_r(x)]^{m_r}$ as its minimal polynomial, where $p_i(x)$'s are distinct monic irreducible polynomials and $m_i \geq 1$. Prove that the matrix of T in the Jordan Normal Form can be put in the form

$$A = \begin{pmatrix} R_1 & O & O & \cdots & O \\ O & R_2 & O & \cdots & O \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ O & O & O & \cdots & R_t \end{pmatrix}$$

such that

$$R_i = \begin{pmatrix} C(p_i(x)^{e_{i_1}}) & & \\ & C(p_i(x)^{e_{i_2}}) & \\ & & \ddots \\ & & & C(p_i(x)^{e_{i_{t_i}}}) \end{pmatrix}$$

with $m_i = e_{i_1} \geq e_{i_2} \geq \dots \geq e_{i_{t_i}}$ and where for any polynomial $g(x)$, $C(g(x)^m)$ denotes the Jordan matrix of $[g(x)]^m$ relative to $g(x)$.

(Matrix A is also called a matrix in *Rational Canonical Form*).

[Hint. If $q_1(x), \dots, q_r(x)$ are the elementary divisors of T , then each of them is a power of some $p_i(x)$. Group together those $q_j(x)$ which are powers of same $p_i(x)$.]

3. If $T \in A(V)$ has all its characteristic roots in F , show that the matrix of F in Jordan Normal Form, is triangular in the sense that all its entries below the diagonal are zero.

4. Give a proof, using matrix computation that if A is a triangular $n \times n$ matrix with entries $\lambda_1, \lambda_2, \dots, \lambda_n$ on the diagonal, then

$$(A - \lambda_1 I') (A - \lambda_2 I') \dots (A - \lambda_n I') = O$$

where I' is the $n \times n$ identity matrix.

5. Find the Jordan Normal Form of the matrix representing T in each of the following.

- (i) $T \in A(\mathbb{Q}^2)$ having elementary divisors $x-2, x+3$.
- (ii) $T \in A(\mathbb{Q}^2)$ having elementary divisor x^2+2x-1 .
- (iii) $T \in A(\mathbb{Q}^3)$ having minimal polynomial, x^3+x+1 .
- (iv) $T \in A(\mathbb{R}^4)$ having elementary divisors

$$x^2 + \pi, x^2 + x + 1.$$

- (v) $T \in A(\mathbb{C}^6)$ having minimal polynomial $x^6 - 1$.

6. Write down all 4×4 matrices in Jordan Normal Form over $\mathbb{R}$ which have 1, 2, 3, as their eigen values.

7. Find Jordan Normal Form of each of the following matrices.

$$\begin{array}{ll} \text{(i)} \begin{pmatrix} -1 & 1 \\ 0 & 1 \end{pmatrix} & \text{(ii)} \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \\ \text{(iii)} \begin{pmatrix} 2 & -1 & -1 \\ -4 & -1 & 1 \\ 4 & 3 & 1 \end{pmatrix} & \text{(iv)} \begin{pmatrix} 2 & 0 & 0 \\ -4 & -2 & 0 \\ 4 & 4 & 2 \end{pmatrix} \end{array}$$

8. Write down all 6×6 matrices over $\mathbb{Q}$ which are in Jordan Normal Form such that

- (i) Each of them has $x^3(x^2+1)$ as its minimal polynomial.
- (ii) Each of them has $(x^2+1)(x+1)(x+2)$ as its minimal polynomial.

(iii) Each of them has $(x+1)^2(x+2)^3(x+1)$ as its minimal polynomial.

(iv) Each of them is nilpotent of index less than or equal to 4.

5. Characteristic Polynomials

Definition 16.60. Let $A=(a_{ij})$ be any $n \times n$ matrix over a field F_1 then the determinant of the matrix

$$xI' - A = \begin{pmatrix} x - a_{11} & -a_{12} & \cdots & \cdots & -a_{1n} \\ -a_{21} & x - a_{22} & \cdots & \cdots & -a_{2n} \\ \cdots & \cdots & \ddots & \cdots & \cdots \\ -a_{n1} & -a_{n2} & \cdots & \cdots & x - a_{nn} \end{pmatrix}$$

is called the *characteristic polynomial* of A . (Here I' denotes the $n \times n$ identity matrix.)

Characteristic polynomial of A is a monic polynomial of degree n ; indeed it is of the form

$$x^n - x^{n-1}(\sum a_{ii}) + \cdots + (-1)^n \det A.$$

EXAMPLE 26. Let $A = \begin{pmatrix} 0 & 1 \\ 2 & 3 \end{pmatrix}$. Then

$$xI' - A = \begin{pmatrix} x-0 & -1 \\ -2 & x-3 \end{pmatrix}$$

$$\det(xI' - A) = x(x-3) - 2 = x^2 - 3x - 2$$

Thus $x^2 - 3x - 2$ is the characteristic polynomial of A . It can be easily verified that A is a root of its characteristic polynomial.

Lemma 16.61. Any two similar matrices have same characteristic polynomial.

Proof: Let $A, B \in M_n(F)$ and $B = P^{-1}AP$ for some non-singular matrix P in $M_n(F)$. Then

$$\begin{aligned} \det(xI' - B) &= \det(xP^{-1}I'P - P^{-1}AP) \\ &= \det[P^{-1}(xI' - A)P] \\ &= \det P^{-1} \det(xI' - A) \det P \\ &= \det(xI' - A), \text{ since } \det P \cdot \det P^{-1} = 1. \end{aligned}$$

Hence A and B have same characteristic polynomial. ■

Definition 16.62. Let $T \in A(V)$ and A be a matrix of T relative to an ordered basis of V . Then the characteristic polynomial of A , is also called the *characteristic polynomial* of T .

Each linear transformation admits unique characteristic polynomial. For $T \in A(V)$, if A and A' are matrices of T relative to two ordered bases B and B' respectively, then A and A' are similar (Corollary 12.40). Hence by Lemma 16.61, A and A' have same characteristic polynomial.

Main purpose of this section is to prove that every linear transformation T is a root of its characteristic polynomial and eigen

value of T is a root of its characteristic polynomial. We shall prove this firstly by using Jordan Normal Form of a matrix and later on give an independent proof. We shall presume that the readers are acquainted with determinants over complex numbers. The definition and some of the basic properties of determinants over complex number can be easily extended to determinants over general field, we shall use these results without proving them. For example any $A \in M_n(F)$ is invertible (i.e. non-singular) if and only if $\det A \neq 0$. Thus a linear transformation T on V is *non-singular* or *singular* according as its associated matrix A has $\det A$ non-zero or zero.

Theorem 16.63. Let $T \in A(V)$. Then $c \in F$ is an eigen value of T if and only if $\det(cI' - A) = 0$, where A is a matrix of T relative to some ordered basis of V .

Proof: Now $cI' - A$ is a matrix of $cI - T$ and c is an eigen value of T if and only if for some non-zero $v \in V$, $T(v) = cv \Leftrightarrow (cI - T)v = 0 \Leftrightarrow cI - T$ is singular. But we know that a linear transformation is singular if its matrix has determinant zero. Consequently c is an eigen value of T if and only if $\det(cI' - A) = 0$. ■

Lemma 16.64. For any monic polynomial $f(x) = x^n - \alpha_{n-1}x^{n-1} - \dots - \alpha_0$ over a field F , the characteristic polynomial of the companion matrix of $f(x)$ is $f(x)$ itself.

Proof: We prove the result by induction on the degree of the polynomial. If $n=1$, then $f(x) = x - \alpha_0$. The companion matrix of $f(x)$ is $A = (\alpha_0)$ and characteristic polynomial of A equals

$$\det(xI' - A) = x - \alpha_0 = f(x).$$

Hence the result holds for $n=1$. To apply induction let the result hold for all monic polynomials of degree $< n$. Now the companion matrix of $f(x)$ is

$$A = \begin{pmatrix} 0 & 0 & \dots & 0 & \alpha_0 \\ 1 & 0 & \dots & 0 & \alpha_1 \\ 0 & 1 & \dots & 0 & \alpha_2 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 & \alpha_{n-1} \end{pmatrix}$$

The characteristic polynomial of A is

$$\det(xI' - A) = |xI' - A| = \begin{vmatrix} x & 0 & 0 & \dots & 0 & -\alpha_0 \\ -1 & x & 0 & \dots & 0 & -\alpha_1 \\ 0 & -1 & x & \dots & 0 & -\alpha_2 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & -1 & x - \alpha_{n-1} \end{vmatrix}$$

Expanding by first column, we get

$$\begin{aligned}
 |xI' - A| &= x \begin{vmatrix} x & 0 & \dots & 0 & -\alpha_1 \\ -1 & x & \dots & 0 & -\alpha_2 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & -1 & x - \alpha_{n-1} \end{vmatrix} \\
 &+ \begin{vmatrix} 0 & 0 & 0 & \dots & 0 & -\alpha_0 \\ -1 & x & 0 & \dots & 0 & -\alpha_2 \\ 0 & -1 & x & \dots & 0 & -\alpha_3 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & -1 & x - \alpha_{n-1} \end{vmatrix} \\
 &= xB + C \quad \dots (1)
 \end{aligned}$$

where

$$B = \begin{vmatrix} x & 0 & \dots & 0 & -\alpha_1 \\ -1 & x & \dots & 0 & -\alpha_2 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & -1 & x - \alpha_{n-1} \end{vmatrix} \text{ and}$$

$$C = \begin{vmatrix} 0 & 0 & 0 & \dots & 0 & -\alpha_0 \\ -1 & x & 0 & \dots & 0 & -\alpha_2 \\ 0 & -1 & x & \dots & 0 & -\alpha_3 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & -1 & x - \alpha_{n-1} \end{vmatrix}.$$

Clearly B is the characteristic polynomial of the companion matrix of the polynomial $x^{n-1} - \alpha_{n-1}x^{n-2} - \dots - \alpha_1$.

As this polynomial is of degree $n-1 < n$, so by induction hypothesis

$$B = x^{n-1} - \alpha_{n-1}x^{n-2} - \dots - \alpha_1 \quad \dots (2)$$

Expanding C by first row we get

$$C = (-1)^{n-2}(-\alpha_0) \begin{vmatrix} -1 & x & 0 & 0 & \dots & 0 \\ 0 & -1 & x & 0 & \dots & 0 \\ 0 & 0 & -1 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 & \dots & -1 \end{vmatrix}$$

[a determinant with $(n-2)$ rows].

$$\text{Thus } C = (-1)^{n-2}(-\alpha_0)(-1)^{n-2} = -\alpha_0 \quad \dots (3)$$

Hence from (1), (2) and (3) we get

$$\begin{aligned}
 |xI' - A| &= x(x^{n-1} - \alpha_{n-1}x^{n-2} - \dots - \alpha_1) - \alpha_0 \\
 &= x^n - \alpha_{n-1}x^{n-1} - \dots - \alpha_1x - \alpha_0 = f(x).
 \end{aligned}$$

Thus the result holds for $\deg f(x) = n$. Hence the result holds for all n . ■

Lemma 16.65. Given any polynomial $g(x) = [f(x)]^m$ where $f(x)$ is a monic polynomial, the characteristic polynomial of the Jordan mat-

rix of $g(x)$ relative to $f(x)$ is $g(x)$ itself.

Proof. Let $\deg f(x) = k$. Then the Jordan matrix of $g(x)$ relative to $f(x)$ is the $km \times km$ matrix

$$A = \begin{pmatrix} C & N & O & O & \dots & O \\ O & C & N & O & \dots & O \\ O & O & C & N & \dots & O \\ \dots & \dots & \dots & \dots & \dots & \dots \\ O & O & O & O & \dots & C \end{pmatrix}$$

where C is the companion matrix of $f(x)$ and N is $k \times k$ matrix

$$\begin{pmatrix} 0 & 0 & \dots & 0 & 1 \\ 0 & 0 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 0 & 0 \end{pmatrix}$$

The characteristic polynomial of A is

$$|xI' - A| = \begin{vmatrix} xI'' - C & -N & O & \dots & O \\ O & xI'' - C & -N & \dots & O \\ \dots & \dots & \dots & \dots & \dots \\ O & O & O & \dots & O \end{vmatrix}$$

where I', I'' are $km \times km$ identity matrix and $k \times k$ identity matrix respectively. Using the block expansion of determinant, we see that

$$|xI' - A| = |xI'' - C|^m,$$

since in the block determinant all entries below the main diagonal are zero.

Now by Lemma 16.64, $|xI'' - C| = f(x)$.

Hence $|xI' - A| = [f(x)]^m = g(x)$.

This proves the lemma. ■

Following lemma is easy to prove.

Lemma 16.66. For any matrix A of the form

$$\begin{pmatrix} A_1 & O & O & \dots & O \\ O & A_2 & O & \dots & O \\ \dots & \dots & \dots & \dots & \dots \\ O & O & O & \dots & A_t \end{pmatrix}$$

where A_i 's are also matrices, the characteristic polynomial of A is equal to the product of characteristic polynomials of A_i 's ($i=1, 2, \dots, t$). ■

Theorem 16.67. (Cayley-Hamilton Theorem). Any linear transformation on a vector space V is a root of its characteristic polynomial.

Proof: Let $m(x)$ be the minimal polynomial of T . Thus to show that T is a root of its characteristic polynomial $f(x)$ it is enough to show that $m(x)$ divides $f(x)$.

Now let $q_1(x), q_2(x), \dots, q_t(x)$ be the elementary divisors of T

(Definition 16.55). Theorem 16.54 yields that the minimal polynomial of T is LCM of these $q_i(x)$'s.

Now the Jordan Normal Form of the matrix of T is

$$A = \begin{pmatrix} J_1 & O & O & \cdots & O \\ O & J_2 & O & \cdots & O \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ O & O & O & \cdots & J_t \end{pmatrix}$$

where each J_i is the Jordan Canonical matrix of $q_i(x)$. Now by Lemma 16.66, the characteristic polynomial $f(x)$ of A (hence of T) is the product of characteristic polynomials of J_i ($i=1, 2, \dots, t$). By Lemma 16.65 the characteristic polynomial of each J_i is $q_i(x)$. Hence

$$f(x) = q_1(x)q_2(x)\cdots q_t(x). \quad \dots(4)$$

Since the minimal polynomial $m(x)$ of T is LCM of $q_i(x)$'s ($i=1, 2, \dots, t$), equation (4) yields that $m(x)$ divides $f(x)$.

Hence $f(T)=0$. This proves the result. ■

REMARK. The above proof indeed shows that the characteristic polynomial of a linear transformation T is equal to the product of its elementary divisors.

Alternative Proof of Theorem 16.67

Let A be the matrix of T relative to some ordered basis $\mathbf{B}$ of V . Then the characteristic polynomial of T is $|xI' - A|$.

Suppose $|xI' - A| = a_0 + a_1x + \dots + a_{n-1}x^{n-1} + x^n$ with a_i 's $\in F$. Let $B(x)$ be the adjoint matrix of the matrix $xI' - A$. Since the elements of $B(x)$ are the co-factors of the elements of matrix $xI' - A$, they are polynomials in x of degree $\leq n-1$. This yields that

$$B(x) = B_0 + B_1x + \dots + B_{n-1}x^{n-1}$$

where for each $i=0, 1, 2, \dots, n-1$; B_i is an $n \times n$ matrix over F . In other words B_i 's are independent of x .

By a well-known property of adjoint of matrices, we get

$$[\text{Adj}(xI' - A)](xI' - A) = [B(x)](xI' - A) = |xI' - A| I'.$$

Thus

$$(B_0 + B_1x + \dots + B_{n-1}x^{n-1})(xI' - A) = (a_0 + a_1x + \dots + a_{n-1}x^{n-1} + x^n)I'.$$

By comparing the coefficients of same powers of x on both sides, we get

$$\begin{aligned} -B_0A &= a_0I' \\ -B_1A + B_0 &= a_1I' \\ &\dots\dots\dots \\ &\dots\dots\dots \\ -B_{n-1}A + B_{n-2} &= a_{n-1}I' \\ B_{n-1} &= I' \end{aligned}$$

Multiplying these matrix equations respectively by $I', A, A^2, \dots$

A^{n-1} and A^n , we get

$$\begin{aligned} -B_0 A &= a_0 I' \\ -B_1 A^2 + B_0 A &= a_1 A \\ &\dots\dots\dots \\ -B_{n-1} A^n + B_{n-2} A^{n-1} &= a_{n-1} A^{n-1} \\ B_{n-1} A^n &= A^n \end{aligned}$$

Adding the above matrix equations, we obtain

$$a_0 I' + a_1 A + \dots + a_{n-1} A^{n-1} + A^n = O.$$

This in turn yields that

$$a_0 I + a_1 T + \dots + a_{n-1} T^{n-1} + T^n = O$$

Hence T is a root of its characteristic polynomial. ■

WORKED-OUT EXERCISES

Exercise 1. *Definition.* A matrix (or linear transformation) is said to be diagonalizable if it is similar to a diagonal matrix (has a basis relative to which its matrix is a diagonal matrix).

If T is in $A(V)$, then T is diagonalizable (if all the roots of its minimal polynomial are in F) if and only if whenever $(T - \lambda I)^{m_v} = 0$ for $v \in V$ and $\lambda \in F$ then $(T - \lambda I)v = 0$.

Solution. Let $m(x)$ be the minimal polynomial of T . Since all its roots are in F , we can write $m(x) = (x - \lambda_1)^{n_1} (x - \lambda_2)^{n_2} \dots (x - \lambda_k)^{n_k}$ where λ_i 's are distinct elements of F and for each $i = 1, 2, \dots, k$; n_i are integers ≥ 1 .

Hence $V = V_1 \oplus V_2 \oplus \dots \oplus V_k$ where each V_i is a T -invariant subspace of V such that the minimal polynomial of T_i , the restriction of T on V_i , is $(x - \lambda_i)^{n_i}$. There exists a basis of V_i relative to which matrix A of T is of the form

$$\begin{pmatrix} J_1 & O & O & \dots & O \\ O & J_2 & O & \dots & O \\ \dots & \dots & \dots & \dots & \dots \\ O & O & O & \dots & J_n \end{pmatrix}$$

where J_i is Jordan matrix of $(x - \lambda_i)^{n_i}$ relative to $x - \lambda_i$. Since for each $v_i \in V_i$, $(T_i - \lambda_i I)^{n_i}(v_i) = 0$, we get $(T - \lambda_i I)v_i = 0$ by hypothesis. Hence the minimal polynomial of T_i on V_i is $x - \lambda_i$. In other words $n_i = 1$ $\forall i = 1, 2, \dots, k$.

Hence for all $i = 1, 2, \dots, k$; J_i is a diagonal matrix, each diagonal element being λ_i . As a consequence A is itself a diagonal matrix. Thus T is diagonalizable.

Conversely suppose T is diagonalizable. There exists an ordered basis $(e_1, e_2, \dots, e_n)$ of V such that $T(e_i) = \lambda_i e_i$ $\forall i = 1, 2, \dots, n$. If $v = 0$ then the result holds trivially.

Let $v \neq 0 \in V$ be such that $[(T - \lambda I)^m]v = 0$ for some $\lambda \in F$, $m \in \mathbb{N}$.

$$\begin{aligned} \text{Now } v &= \sum_{i=1}^n \mu_i e_i, \mu_i \in F \\ &= \sum_{j=1}^t \mu_{i_j} e_{i_j} \end{aligned}$$

where $\{e_{i_1}, e_{i_2}, \dots, e_{i_t}\} \subseteq \{e_1, e_2, \dots, e_n\}$ and each μ_{i_j} is a non-zero element of F .

$$\text{Further } [(T - \lambda I)^m]v = 0 \Rightarrow \sum_{j=1}^t \mu_{i_j} (T - \lambda I)^m e_{i_j} = 0.$$

Again $T(e_i) = \lambda_i e_i$ gives that $T^2(e_i) = \lambda_i T(e_i) = \lambda_i^2 e_i$. Proceeding in this manner it can be shown that for any positive integer r , $T^r(e_i) = \lambda_i^r e_i$.

Hence $(T - \lambda I)^m e_{i_j} = (\lambda_{i_j} - \lambda)^m e_{i_j}$ for all $j = 1, 2, \dots, t$, and $\{\lambda_{i_1}, \lambda_{i_2}, \dots, \lambda_{i_t}\} \subseteq \{\lambda_1, \lambda_2, \dots, \lambda_n\}$.

$$\begin{aligned} \text{Hence } [(T - \lambda I)^m]v = 0 &\Rightarrow \sum_{j=1}^t \mu_{i_j} (\lambda_{i_j} - \lambda)^m e_{i_j} = 0 \\ &\Rightarrow \mu_{i_j} (\lambda - \lambda_{i_j})^m = 0 \forall j \text{ as } e_{i_1}, e_{i_2}, \dots, e_{i_t} \text{ are L.I. vectors over } F. \\ \text{Since each } \mu_{i_j} &\neq 0, \text{ this in turn implies that } \lambda_{i_j} = \lambda \forall j = 1, 2, \dots, t. \\ \text{Consequently } (T - \lambda I)v &= \sum_{j=1}^t \mu_{i_j} (T - \lambda I)e_{i_j} = \sum_{j=1}^t \mu_{i_j} (\lambda_{i_j} - \lambda) e_{i_j} \\ &= 0. \end{aligned}$$

Exercise 2. Prove that if a linear transformation T on V satisfies $T^2 = T$, then T is diagonalizable. Further if S and $T \in A(V)$ such that $S^2 = S$ and $T^2 = T$ then S and T are similar if and only if they have the same rank.

Solution. Note that in the solution of Exercise 1, we required for first part the condition $[(T - \lambda I)^m]v = 0$ implies $(T - \lambda I)v = 0$ for all $v \in V$ and for all roots λ of minimal polynomial $m(x)$ of T , provided all the roots of $m(x)$ lie in F . In this Exercise $T^2 = T$ gives that $m(x) = x^2 - x$. So the only roots of $m(x)$ are 1 and 0, which obviously are in F .

Now $[(T - 0I)^m]v = 0 \Rightarrow T^m(v) = 0$ for any $m \in \mathbb{N}$. If $m = 1$, this gives $T(v) = 0$.

If $m \geq 2$ then by the relation $T^2 = T$, $T^m = T$. Once again we get that $[(T - 0I)^m]v = 0 \Rightarrow T(v) = 0$.

Again note that $(T - I)^2 = T^2 - 2T + I = I - T$. i.e. $(I - T)^2 = I - T$. As before this implies that for all $m \geq 2$, $(I - T)^m = I - T$. Hence $[(T - I)^m]v = 0 \Rightarrow [(I - T)^m]v = 0 \Rightarrow [(I - T)]v = 0 \Rightarrow [(T - I)]v = 0$.

Thus the condition of first part of Exercise 1 given above, is satisfied. Hence T is diagonalizable.

Now suppose S and T are similar then there exists a non-singular $U \in A(V)$ such that $S = UTU^{-1}$.

Then $r(S) = r(UTU^{-1}) = r(U^{-1}UT) = r(T)$.

Conversely let $r(S) = r(T) = k$.

Since $S^2 = S$ and $T^2 = T$, both S and T are diagonalizable, there exist ordered basis $\{e_1, e_2, \dots, e_n\}$ and $\{f_1, f_2, \dots, f_n\}$ of V such that

$$S(e_i) = e_i \quad \forall 1 \leq i \leq k \\ = 0 \quad \forall k < i \leq n,$$

and

$$T(f_i) = f_i \quad \forall 1 \leq i \leq k \\ = 0 \quad \forall k < i \leq n.$$

Define $U \in A(V)$ by $U(e_i) = f_i \quad \forall i = 1, 2, \dots, n$. Then U is non-singular and $US(e_i) = U(e_i) = f_i = T(f_i) = TU(e_i) \quad \forall i = 1, 2, \dots, k$. For $i > k$, $US(e_i) = 0 = T(f_i) = TU(e_i)$. This gives, that $US = TU \Rightarrow T = USU^{-1}$. Hence T and S are similar.

PROBLEMS

1. Show that the characteristic polynomial of a linear transformation equals its minimal polynomial if and only if the linear transformation is cyclic.

2. For each of the following polynomials over $\mathbb{Q}$ write its Jordan Canonical matrix and determine its characteristic polynomial

$$(x^2 + 1)^2, (x + 1)^3, (x^3 + 2x + 1)^2, x^3 - 2.$$

3. Show that $n \times n$ matrix A is similar to a diagonal matrix if and only if its minimal polynomial $m(x)$ is equal to $(x - \alpha_1)(x - \alpha_2) \dots (x - \alpha_m)$ for some distinct elements $\alpha_1, \alpha_2, \dots, \alpha_m$ of F .

[Hint. Use Jordan Normal Form].

4. Show by direct computations that the none of the following matrices over $\mathbb{Q}$ is similar to a diagonal matrix.

$$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \quad \begin{pmatrix} 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{pmatrix}, \quad \begin{pmatrix} 0 & -2 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

5. Show that the matrices $\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ and $\begin{pmatrix} 0 & -2 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}$

as matrices over $\mathbb{C}$ are similar to diagonal matrices.

6. Find the characteristic polynomial of each of the following matrices over $\mathbb{Q}$

$$\begin{pmatrix} 1 & 0 & 0 \\ 2 & 3 & 0 \\ -2 & -2 & 1 \end{pmatrix}, \quad \begin{pmatrix} 2 & 3 & 0 \\ 0 & 2 & 3 \\ 3 & 2 & 0 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \\ 3 & 1 & 2 \end{pmatrix}, \quad \begin{pmatrix} -4 & 1 \\ 0 & 4 \end{pmatrix}.$$

7. If $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in M_2(\mathbf{R})$ with either $b \neq 0$ or $c \neq 0$, show that the characteristic polynomial of A is $f(x) = x^2 - x(a+d) + ad - bc$ and $f(x)$ is also the minimal polynomial of A . Deduce that A is similar to a diagonal matrix over $\mathbf{R}$ if and only if $(a+d)^2 - 4(ad - bc) > 0$.

APPENDIX

1. FUNDAMENTAL THEOREM OF FINITELY GENERATED ABELIAN GROUPS

In Chapter 4, Fundamental Theorem of Finite Abelian Groups was proved (Theorem 4.26). Here the decomposition part of that theorem is extended to finitely generated Abelian groups.

Definition. An Abelian group G is said to be *finitely generated*, if there exists a finite subset S of G such that $G = \langle S \rangle$. S is called a *generating set* of G .

For example every cyclic group is finitely generated since any subset which has only one element namely a generator of the cyclic group, works as a generating set. Klein's 4-group $\{e, a, b, ab \mid a^2 = b^2 = e, ab = ba\}$ is a finitely generated Abelian group, generated by $\{a, b\}$. In fact every finite Abelian group is finitely generated. Consider M , the additive group of 2×2 matrices over integers. This

is a finitely generated Abelian group. The set $S = \left\{ \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \right\}$ is one of its generating sets.

M is an infinite group.

Theorem. (Fundamental Theorem of Finitely Generated Abelian Groups). Every finitely generated Abelian group is direct product of a finite number of its cyclic subgroups.

Proof: Let G be a finitely generated Abelian group. Among all its generating sets, choose those which have got smallest number of elements. We shall call each of these generating sets, a *minimal generating set*. The number of elements in every minimal generating set will be termed as *rank* of G .

We apply induction on rank of G . In case rank $G=1$, G itself is a cyclic group and the theorem holds trivially.

Suppose now that rank $G=k>1$ and the theorem is true for all finitely generated Abelian groups of rank less than k .

If $\{a_1, a_2, \dots, a_n\}$ is a minimal generating set of G such that $a_1^{r_1} a_2^{r_2} \dots a_k^{r_k} = e \Rightarrow a_i^{r_i} = e$ for all $i=1, 2, \dots, k$ and for all integers

$r_1, r_2, \dots, r_k$. Then, if $H_i = \langle a_i \rangle$, $H_1, H_2, \dots, H_k$ is direct product of cyclic subgroups H_i and, since $\{a_1, a_2, \dots, a_k\}$ generates G , $G = H_1 \times H_2 \times \dots \times H_k$, we are done.

Therefore, let us assume that for every given minimal generating set $\{a_1, a_2, \dots, a_k\}$ there exist integers $\alpha_1, \alpha_2, \dots, \alpha_k$ such that $a_1^{\alpha_1} a_2^{\alpha_2} \dots a_k^{\alpha_k} = e$ with at least one $a_i^{\alpha_i} \neq e$. We say that α_i occurs as an index in such a relation (as index of i th generator). We can choose $\alpha_i > 0$. Otherwise we replace $a_1^{\alpha_1} a_2^{\alpha_2} \dots a_k^{\alpha_k} = e$ by the relation $a_1^{-\alpha_1} a_2^{-\alpha_2} \dots a_k^{-\alpha_k} = e^{-1} = e$. Now $-\alpha_i > 0$ such that $a_i^{-\alpha_i} \neq e$. In other words if $\alpha_i < 0$ then $-\alpha_i > 0$ occurs as an index in a relation satisfied by $\{a_1, a_2, \dots, a_k\}$.

Among all possible such relations for all minimal generating sets there is a smallest positive integer occurring as an index. Let this integer be t and let the generating set for which it occurs be $\{c_1, c_2, \dots, c_k\}$. Suppose t occurs as an index of c_j so $c_j^t \neq e$ and we have $c_1^{l_1} c_2^{l_2} \dots c_k^{l_k} = e$. For convenience sake, we manipulate things in such a manner that t occurs as index of first generator. If $j \neq 1$, we proceed as under.

Put $b_i = c_i$ and $l_i = n_i$ for all i except $i=1$ and $i=j$. Also put $b_1 = c_j$, $b_j = c_1$ and $l_i = n_j$, $l_j = n_1 = t$.

Thus we have a relation $b_1^{n_1} b_2^{n_2} \dots b_k^{n_k} = e$ and $b_1^{n_1} = c_j^{n_1} = c_j^{l_j} = c_j^t \neq e$. Further if for any generating set $\{\lambda_1, \lambda_2, \dots, \lambda_k\}$, $\lambda_1^{h_1} \lambda_2^{h_2} \lambda_3^{h_3} \dots \lambda_k^{h_k} = e$ with at least one $\lambda_i^{h_i} \neq e$, $n_i \leq h_j$ for each b_j occurring as index in the relation $\lambda_1^{h_1} \lambda_2^{h_2} \dots \lambda_k^{h_k} = e$ (A)

We prove the main Theorem by means of several claims.

Claim I. If $b_i^{n_i} \neq e$ and b_i is of finite order then $n_i \leq o(b_i)$.

Suppose $q_i = o(b_i) < n_i$.

Now $n_i = p_i q_i + u_i$ with $0 \leq u_i < q_i$.

so $b_i^{n_i} = (b_i^{q_i})^{p_i} b_i^{u_i} = b_i^{u_i}$.

Thus $b_1^{n_1} b_2^{n_2} \dots b_i^{n_i} \dots b_k^{n_k} = b_1^{n_1} b_2^{n_2} \dots b_i^{u_i} \dots b_k^{n_k} = e$. Since $b_i^{u_i} = b_i^{n_i} \neq e$, we have a relation in which u_i occurs as index.

We can't have $0 < u_i < q_i$, since in that case $u_i \leq n_1$ and statement (A) will be violated.

Hence $u_i = 0$. This in turn implies that $b_i^{n_i} = b_i^{u_i} = e$, a contradiction to hypothesis of the claim.

Consequently our supposition is wrong. In other words if b_i is of finite order q_i , then $n_i \leq q_i$. In particular as $b_1^{n_1} \neq e$, if b_1 is of finite order then $n_1 \leq o(b_1)$ (B)

Claim II. If $b_1^{m_1} b_2^{m_2} \dots b_k^{m_k} = e$ then n_1/m_1 .

Let $m_1 = n_1 g + l$ with $0 \leq l < n_1$.

Now $b_1^{m_1} b_2^{m_2} \dots b_k^{m_k} = e$

$$\Rightarrow b_1^{n_1 g + l} b_2^{m_2} \dots b_k^{m_k} = e$$

$$\Rightarrow b_1^l b_2^{m_2 - n_2 g} \dots b_k^{m_k - n_k g} = e \text{ as } b_1^{n_1 g} = b_2^{-n_2 g} \dots b_k^{-n_k g}.$$

Suppose $b_1^l \neq e$ then l appears as index in a relation satisfied by $\{b_1, b_2, \dots, b_k\}$. This in view of statement (A) forces $l=0$, a contradiction to the fact $b_1^l \neq e$.

Hence $b_1^l = e$.

If $l \neq 0$ then this implies that $0(b_i) \leq l < n_1$. But then statement (B) is violated.

Consequently $l=0$ and $m_1 = n_1 g \Rightarrow n_1/m_1$.

Claim III. If for any $i=2, 3, \dots, k$; $b_i^{n_i} \neq e$ then n_1/n_i .

Suppose that $n_i = n_1 \alpha_i + v_i$, $0 \leq v_i < n_1$. Take $b_1' = b_1 b_i^{\alpha_i}$, $b_2' = b_2, \dots$, $b_k' = b_k$.

For any $x \in G$,

$$\begin{aligned} x &= b_1^{\beta_1} b_2^{\beta_2} \dots b_k^{\beta_k}; \beta_k \in \mathbb{Z} \\ &= (b_1 b_i^{\alpha_i})^{\beta_1} b_2^{\beta_2} \dots b_i^{\beta_i - \alpha_i \beta_1} \dots b_k^{\beta_k} \\ &= (b_1')^{\beta_1} (b_2')^{\beta_2} \dots (b_i')^{\beta_i - \alpha_i \beta_1} \dots (b_k')^{\beta_k}. \end{aligned}$$

Hence $b_1', b_2', b_3', \dots, b_k'$ generate G .

More over $(b_1')^{n_1} (b_2')^{n_2} \dots (b_i')^{v_i} (b_k')^{n_k}$

$$= b_1^{n_1} b_2^{n_2} \dots b_i^{v_i + n_1 \alpha_i} \dots b_k^{n_k}$$

$$= b_1^{n_1} b_2^{n_2} \dots b_i^{n_i} \dots b_k^{n_k} = e.$$

$(b_i')^{v_i} \neq e$ violates statement (A)

Now

Hence $(b_i')^{v_i} = e$, i.e. $b_i^{v_i} = e$.

If $v_i \neq 0$ then $0(b_i) \leq v_i < n_1$. This is against claim I.

Therefore $v_i = 0$ and so n_1/v_i .

Finally put $b_1^* = b_1 b_2^{\alpha_2} b_3^{\alpha_3} \dots b_k^{\alpha_k}$, where α_i satisfy the following conditions.

If $b_i^{n_i} \neq e$ then take $\alpha_i = \frac{n_i}{n_1}$ and if $b_i^{n_i} = e$ then take $\alpha_i = 0$.

$$\begin{aligned} \text{Let } x \in G, \quad x &= b_1^{\delta_1} b_2^{\delta_2} \dots b_k^{\delta_k}, \delta_i \in \mathbb{Z} \\ &= (b_1^* b_2^{-\alpha_2} \dots b_k^{-\alpha_k})^{\delta_1} b_2^{\delta_2} \dots b_k^{\delta_k} \\ &= (b_1^*)^{\delta_1} b_2^{\delta_2 - \alpha_2 \delta_1} \dots b_k^{\delta_k - \alpha_k \delta_1} \end{aligned}$$

Hence $\{b_1^*, b_2, b_3, \dots, b_k\}$ generates G .

$$\text{Also } (b_1^*)^{n_1} = b_1^{n_1} b_2^{\alpha_2 n_1} b_3^{\alpha_3 n_1} \dots b_k^{\alpha_k n_1} \quad (i)$$

$$= b_1^{n_1} b_2^{n_2} \dots b_k^{n_k} = e$$

$$(b_1^*)^{s_1} b_2^{s_2} \dots b_k^{s_k} = e$$

$$b_1^{s_1} b_2^{s_2 + \alpha_2 s_1} \dots b_k^{s_k + \alpha_k s_1} = e.$$

By claim II, n_1/s_1 so $s_1 = n_1 t_1$, $t_1 \in \mathbb{Z}$

Hence $(b_1^*)^{s_1} = (b_1^*)^{n_1 t_1} = [(b_1^*)^{n_1}]^{t_1} = e$ by (1).

Now let A_1 be cyclic subgroup of G generated by b_1^* and A_2 the subgroup generated by $b_2, b_3, \dots, b_k$.

Then clearly $G = A_1 A_2$. Suppose $x \in A_1 \cap A_2$, then $x = (b_1^*)^a = b_2^{a_2} b_3^{a_3} \dots b_k^{a_k}$. This implies $(b_1^*)^a = b_2^{-a_2} b_3^{-a_3} \dots b_k^{-a_k} = e$.

As proved above this implies $(b_1^*)^a = e$ i.e. $x = e$, hence $A_1 \cap A_2 = (e)$. In other words $G = A_1 \times A_2$.

But now A_2 is generated by $k-1$ or less than $k-1$ generators, so by induction hypothesis A_2 is direct product of cyclic subgroups. Consequently G itself is direct product of cyclic subgroups.

2. SOME COUNTER-EXAMPLES

In the following we shall discuss three counter-examples which could not find place in text proper. In Theorem 10.4 we proved that in an integral domain with unity the elements a and b are associates if and only if $a \mid b$ and $b \mid a$. In Section 1 we give two elements a and b of ring R (which is not an integral domain) such that $a \mid b$ and $b \mid a$, yet a and b are not associates. In Section 2 we give two commutators of a group, whose product is not a commutator. Finally in Section 3 we give an example of a PID which is not a Euclidean Domain.

1. Fletcher's Example

Consider the polynomial ring $F[x, y, z]$ in three indeterminates x, y, z over a field F . (Note that we discussed in Chapter 9, polynomial rings in one variable; here $F[x, y, z] = S[y, z]$ where $S = F[x]$ and then $S[y, z] = T[z]$ where $T = S[y]$, thus $F[x, y, z] = T[z]$.) Put $R = \frac{F[x, y, z]}{\langle x - xyz \rangle}$. Let $A = \langle x - xyz \rangle$ and let for any $r \in F[x, y, z]$, $\bar{r} = r + A \in R$. Then $x - xyz = \bar{0} \Rightarrow \bar{x}(1 - yz) = \bar{0}$, but neither $\bar{x} = \bar{0}$ nor $\bar{1} - yz = \bar{0}$, so R is not an integral domain. Put $\bar{a} = \bar{x}$ and $\bar{b} = xy$. Since $b = \overline{xy} = \bar{x}\bar{y} = a\bar{y}$ we get that $a \mid b$. Also $\overline{x - xyz} = \bar{0} \Rightarrow \bar{x} = \overline{xyz} \Rightarrow a = b\bar{z} \Rightarrow b \mid a$.

Suppose that a and b are associates. Then $a = bu$ for some unit u in R . This in turn implies that $x = xyu + (x - xyz)f$ for some $f \in F[x, y, z]$. Further there exists v in R such that $uv = \bar{1}$, i.e., $uv = 1 + (x - xyz)g$ for some $g \in F[x, y, z]$. Thus we can write $u = \alpha + xh$ and $v = \alpha^{-1} + xk$ for some $\alpha (\neq 0) \in F$ and $h, k \in F[x, y, z]$. Now $x = xyu + (x - xyz)f \Rightarrow 1 = yu + (1 - yz)f \Rightarrow 1 = y(\alpha + xh) + (1 - yz)f \Rightarrow 1 - y\alpha = xyh + (1 - yz)f$. Comparing the terms independent of x on both sides of last equation, we obtain $1 - y\alpha = (1 - yz)f(0, y, z)$. Thus $f(0, y, z) = 1$ and so $z = \alpha$. This is absurd! Hence a and b can not be associates.

2. Scot's Example

Let $S = \{0, a, b, c, d, f, g, i, 1\}$. Define $ab=f, ac=g, ad=h, bc=i, 1x=x1=x$ for all $x \in S$ and all other products to be equal to zero. S is a semi-group under the product defined above. Consider R , the semi-group ring of S over $\mathbb{Z}_2 = \{0, 1\}$, the field of integers modulo 2. (Cf. Worked-out Exercise 4, Section 3, Chapter 8).

$$\text{Put } G = \left\{ \begin{pmatrix} 1 & x & y \\ 0 & 1 & x \\ 0 & 0 & 1 \end{pmatrix} \mid x, y \in R \right\}.$$

It can be easily verified that G is a group under matrix multiplication. Note that $\begin{pmatrix} 1 & x & y \\ 0 & 1 & x \\ 0 & 0 & 1 \end{pmatrix}^{-1} = \begin{pmatrix} 1 & -x & x^2 - y \\ 0 & 1 & -x \\ 0 & 0 & 1 \end{pmatrix}$. Firstly we assert

that every commutator of G is of the type $\begin{pmatrix} 1 & 0 & z \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$ where z is a commutator of the ring R (in other words $z = uv - vu$ for some $u, v \in R$).

Let g be a commutator of G . Then $g = ABA^{-1}B^{-1}$ for some $A, B \in G \Rightarrow g = \begin{pmatrix} 1 & x & y \\ 0 & 1 & x \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & u & v \\ 0 & 1 & u \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & -x & x^2 - y \\ 0 & 1 & -x \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & -u & u^2 - v \\ 0 & 1 & -u \\ 0 & 0 & 1 \end{pmatrix}$
 $= \begin{pmatrix} 1 & 0 & xu - ux \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$. Hence our assertion follows.

Now $h = ad = ad - da$ and $i = bc = bc - cb$ are commutators of R . (Recall that $da = cb = 0$). Yet, we claim, that $h + i$ is not a commutator of R . Otherwise, $h + i = (\sum_{j=1}^9 \alpha_j g_j) (\sum_{j=1}^9 \beta_j g_j) - (\sum_{i=1}^9 \beta_i g_i) (\sum_{i=1}^9 \alpha_i g_i)$ for some $\alpha_i, \beta_i \in \mathbb{Z}_2$ and $g_1 = a, g_2 = b, g_3 = c, g_4 = d, g_5 = f, g_6 = g, g_7 = h, g_8 = i$ and $g_9 = 1$. Using the defining relations of semi-group S we obtain $h + i = (\alpha_1 \beta_2 - \alpha_2 \beta_1) f + (\alpha_1 \beta_3 - \alpha_3 \beta_1) g + (\alpha_1 \beta_4 - \alpha_4 \beta_1) h + (\alpha_2 \beta_3 - \alpha_3 \beta_2) i \Rightarrow \alpha_1 \beta_2 - \alpha_2 \beta_1 = 0 = (\alpha_1 \beta_3 - \alpha_3 \beta_1)$ and $\alpha_1 \beta_4 - \alpha_4 \beta_1 = 1 = \alpha_2 \beta_3 - \alpha_3 \beta_2$. Now α_1 can take only two values namely 0 or 1. In case $\alpha_1 = 0$ we get that $-\alpha_4 \beta_1 = 1 \Rightarrow \alpha_4 = \beta_1 = 1 \Rightarrow \alpha_2 = \alpha_3 = 0 \Rightarrow 1 = 0$. This is absurd. In case $\alpha_1 = 1$, we get $\beta_2 = \alpha_2 \beta_1, \beta_3 = \alpha_3 \beta_1$. Thus $\alpha_2 \beta_3 - \alpha_3 \beta_2 = 1 \Rightarrow \alpha_2 (\alpha_3 \beta_1) - \alpha_3 (\alpha_2 \beta_1) = 1 \Rightarrow 0 = 1$, again a contradiction. Consequently our claim is established.

Finally $\begin{pmatrix} 1 & 0 & h \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$ and $\begin{pmatrix} 1 & 0 & i \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$ are commutators of G , for

$$\begin{aligned}
& \begin{pmatrix} 1 & a & 0 \\ 0 & 1 & a \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & d & 0 \\ 0 & 1 & d \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & -a & a^2 \\ 0 & 1 & -a \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & -d & d^2 \\ 0 & 1 & -d \\ 0 & 0 & 1 \end{pmatrix} \\
&= \begin{pmatrix} 1 & 0 & ad-da \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & h \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \text{ and} \\
& \begin{pmatrix} 1 & b & 0 \\ 0 & 1 & b \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & c & 0 \\ 0 & 1 & c \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & -b & b^2 \\ 0 & 1 & -b \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & -c & c^2 \\ 0 & 1 & -c \\ 0 & 0 & 1 \end{pmatrix} \\
&= \begin{pmatrix} 1 & 0 & bc-cb \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & i \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.
\end{aligned}$$

But the product $\begin{pmatrix} 1 & 0 & h \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & i \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & h+i \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$ is not a commutator, otherwise $h+i$ must be a commutator of R , which is not possible by our claim, as proved above.

3. Motzkin's Example

The discussion given below is a slightly varied form of the note written by J.C. Wilson of University of North Carolina at Asheville, in *Mathematic Magazine*, Vol 46, No. 1, 1973.

Consider $R = \left\{ a + \frac{b}{2} (1 + \sqrt{-19}) \mid a, b \in \mathbf{Z} \right\}$. We shall prove that R is a PID but not a Euclidean Domain. This we shall do with the help of several lemmas.

Lemma 1. (i) R is a sub-domain of $\mathbf{C}$. (ii) R has unity. (iii) The only units of R are ± 1 (iv) For all $a, b \in \mathbf{Z}$, $a \pm b\sqrt{-19} \in R$.

Proof: (i) Clearly $0 \in R$, so R is non-empty. Let $\alpha, \beta \in R$. Then $\alpha = a + \frac{b}{2} (1 + \sqrt{-19})$, $\beta = c + \frac{d}{2} (1 + \sqrt{-19})$ for some $a, b, c, d \in \mathbf{Z}$. Now $\alpha - \beta = (a - c) + \frac{(b - d)}{2} (1 + \sqrt{-19}) \in R$; also $\alpha\beta = (ac - 5bd) + \frac{1}{2}(bd + ad + bc)(1 + \sqrt{-19}) \in R$. Thus R is a subring of $\mathbf{C}$. As $\mathbf{C}$ is an integral domain, R is a subdomain of $\mathbf{C}$.

(ii) $1 = 1 + \frac{0}{2} (1 + \sqrt{-19}) \in R$.

(iii) Let $a + \frac{b}{2} (1 + \sqrt{-19})$ be a unit of R . Then there exists $c + \frac{d}{2} (1 + \sqrt{-19}) \in R$ such that

$$\begin{aligned} & \left[a + \frac{b}{2} (1 + \sqrt{-19}) \right] \left[c + \frac{d}{2} (1 + \sqrt{-19}) \right] = 1. \\ \Rightarrow & \left[\left(a + \frac{b}{2} \right) + \frac{b}{2} \sqrt{-19} \right] \left[\left(c + \frac{d}{2} \right) + \frac{d}{2} \sqrt{-19} \right] = 1 \\ \Rightarrow & \left[\left(a + \frac{b}{2} \right) - \frac{b}{2} \sqrt{-19} \right] \left[\left(c + \frac{d}{2} \right) - \frac{d}{2} \sqrt{-19} \right] = 1 \end{aligned}$$

Thus we get $\left[\left(a + \frac{b}{2} \right)^2 + \frac{19b^2}{4} \right] \left[\left(c + \frac{d}{2} \right)^2 + \frac{19d^2}{4} \right] = 1$

i.e., $(a^2 + ab + 5b^2)(c^2 + cd + 5d^2) = 1$

This gives that $a^2 + ab + 5b^2 = c^2 + cd + 5d^2 = 1$

[Note that $a^2 + ab + 5b^2 = \left(a + \frac{b}{2} \right)^2 + \frac{19b^2}{4} \geq 0$].

In other words $\left(a + \frac{b}{2} \right)^2 + \frac{19b^2}{4} = 1$. If $b \neq 0$, then $b^2 \geq 1$,

$\Rightarrow \frac{19b^2}{4} \geq 1$ and we get a contradiction so $b = 0$ and as a consequence

$a^2 = 1$ or $a = \pm 1$. Hence the only units of R are ± 1 .

(iv) $a + b\sqrt{-19} = (a - b) + \frac{2b}{2} (1 + \sqrt{-19})$ and

$a - b\sqrt{-19} = (a + b) - \frac{(-2b)}{2} (1 + \sqrt{-19})$ imply that

$a \pm b\sqrt{-19} \in R$. ■

For any $x \in \mathbf{C}$ we define $N(x)$ to be $|x|^2$. Thus if $x = a + \frac{b}{2} (1 + \sqrt{-19})$ then $N(x) = \left(a + \frac{b}{2} \right)^2 + \frac{19b^2}{4} = a^2 + ab + 5b^2$. So for each $x \in R$, $N(x)$ is a non-negative integer. Also for any $x \in \mathbf{C}$, $N(x) = 0 \Leftrightarrow x = 0$. Observe that for all $x, y \in \mathbf{C}$, $N(xy) = N(x)N(y)$ and if $y \neq 0 \in \mathbf{C}$ then $N(y^{-1}) = [N(y)]^{-1} \in \mathbf{Q}$.

Lemma 2. If for all pairs of non-zero elements α, β in R with $N(\alpha) \geq N(\beta)$ either $\beta \mid \alpha$ or there exist γ, δ in R such that $0 < N(\alpha\gamma - \beta\delta) < N(\beta)$ then R is a PID.

Proof: In view of Lemma 1, it will be sufficient to prove that every non-zero ideal of R is principal. Let I be an ideal of R , different from (0) .

Let $\mathcal{F} = \{n \in \mathbf{Z} \mid n > 0 \text{ and there exists } x \in I \text{ with } N(x) = n\}$. Since $I \neq (0)$, $\mathcal{F}$ is non-empty. Let m be the least element of $\mathcal{F}$ and let $y \in I$ be such that $N(y) = m$. Thus for any $x (\neq 0) \in I$, $N(x) \geq N(y)$. Now for all γ and δ in R , $x\gamma - y\delta \in I$ as I is an ideal of R . Either $x\gamma - y\delta = 0 \forall \gamma, \delta \in R$ or for some $\gamma, \delta \in R$, $N(x\gamma - y\delta) \geq N(y)$. In case $N(x\gamma - y\delta) \geq N(y)$ we get by hypothesis that $y \mid x$ and so $x \in \langle y \rangle$. In the other case $x\gamma - y\delta = 0 \forall \gamma, \delta \in R \Rightarrow x\gamma = y\delta \forall \gamma, \delta \in R$. In parti-

cular $x = \delta y$ i.e., $x \in \langle y \rangle$. So in each case $x \in \langle y \rangle$. In other words $I = \langle y \rangle$. Hence R is a PID. ■

Lemma 3. For all $\alpha, \beta, \gamma, \delta$ in R , $0 < N(\alpha\gamma - \beta\delta) < N(\beta)$ if and only if, $0 < N\left(\frac{\alpha\gamma}{\beta} - \delta\right) < 1$.

Proof: Now $0 < N(\alpha\gamma - \beta\delta) < N(\beta) \Rightarrow \beta \neq 0 \Rightarrow \beta^{-1} \in \mathbf{C}$. Thus $N\left(\frac{\alpha\gamma}{\beta} - \delta\right) = N(\alpha\gamma - \beta\delta) N(\beta^{-1}) = \frac{N(\alpha\gamma - \beta\delta)}{N(\beta)} < 1$. As $N(\alpha\gamma - \beta\delta) > 0$ we get that $0 < N\left(\frac{\alpha\gamma}{\beta} - \delta\right) < 1$. Conversely suppose that $0 < N\left(\frac{\alpha\gamma}{\beta} - \delta\right) < 1$. This gives that $0 < \frac{N(\alpha\gamma - \beta\delta)}{N(\beta)} < 1 \Rightarrow 0 < N(\alpha\gamma - \beta\delta) < N(\beta)$. ■

Theorem 1. R is a PID.

Proof: Let x, y be a pair of non-zero elements of R , such that $y \nmid x$ in R , we shall prove that there exist $\gamma, \delta \in R$ such that $0 < N(x\gamma - y\delta) < N(y)$. Then by invoking Lemma 2, we shall get our result.

Since $x, y \in R \Rightarrow x, y \in \mathbf{Q}(\sqrt{-19})$ and $y \neq 0$, we get that $\frac{x}{y} \in \mathbf{Q}(\sqrt{-19})$. In other words $\frac{x}{y} = \alpha + \beta\sqrt{-19}$, $\alpha, \beta \in \mathbf{Q} \Rightarrow \frac{x}{y} = \frac{p + q\sqrt{-19}}{r}$ with $p, q, r \in \mathbf{Z}$, $r \geq 1$. We can assume without loss of generality that $(p, q, r) = 1$. Obviously $r \neq 1$ otherwise $y \mid x$ in R . Hence $r > 1$.

Four cases arise namely (I) $r \geq 5$ (II) $r = 2$, (III) $r = 3$, (IV) $r = 4$.

Case I. $r \geq 5$.

Since $(p, q, r) = 1$, $\exists a, b, c \in \mathbf{Z}$ such that $ap + bq + cr = 1$. Also $pb - 19aq = rs + t$ with $r, t \in \mathbf{Z}$ and $0 \leq t < r$. Note that either $0 \leq t \leq \frac{r}{2}$ or $\frac{r}{2} < t < r$. In case $\frac{r}{2} < t < r$, we put $t' = t - r$. This implies that $0 < -t' = r - t < \frac{r}{2}$, and $pb - 19aq = rs + t = r(s+1) + t'$ with $-\frac{r}{2} < t'$. Thus we can always write $bp - 19aq = rs + t$ with $r, s \in \mathbf{Z}$ and $|t| \leq \frac{r}{2}$. Now put $z = b + a\sqrt{-19}$ and $w = s - c\sqrt{-19}$. By Lemma 1; $z, w \in R$ and $\left(\frac{x}{y}\right)z - w = \frac{(p + q\sqrt{-19})}{r} (b + a\sqrt{-19}) - (s - c\sqrt{-19}) = \frac{t + \sqrt{-19}}{r}$. As $t \neq -\sqrt{-19}$, $N\left[\left(\frac{x}{y}\right)z - w\right] > 0$.

Also $N \left\{ \left(\frac{x}{y} \right) z - w \right\} = \frac{t^2 + 19}{r^2}$. So if $r > 5$ then $r \geq 6 \Rightarrow \frac{1}{r^2} \leq \frac{1}{36}$.

Then using $\frac{t^2}{r^2} \leq \frac{1}{4}$ we obtain that $N \left\{ \left(\frac{x}{y} \right) z - w \right\} \leq \frac{1}{4} + \frac{19}{36} =$

$$\frac{28}{36} = \frac{7}{9} < 1. \text{ Hence } 0 < N \left\{ \left(\frac{x}{y} \right) z - w \right\} < 1. \text{ Further } r = 5 \Rightarrow t \leq \frac{5}{2}.$$

Since t is an integer, $|t| \leq 2$. This gives that $\frac{t^2 + 19}{r^2} = \frac{t^2 + 19}{25}$

$$\leq \frac{23}{25} < 1. \text{ Hence once again } 0 < N \left\{ \left(\frac{x}{y} \right) z - w \right\} < 1.$$

Case II. $r = 2$.

As $(p, q, r) = 1$; both p and q cannot be even simultaneously. Let both of them be odd. Put $p = 2m + 1$, $q = 2n + 1$ for some $m, n \in \mathbb{Z}$.

$$\text{Then } \frac{x}{y} = \frac{p + q\sqrt{-19}}{2} = \frac{(2m+1) + (2n+1)\sqrt{-19}}{2} = (m + n\sqrt{-19})$$

$+ \left(\frac{1 + \sqrt{-19}}{2} \right) \in R$. This forces y to be a factor of x which is

absurd. So one of p and q must be even and the other, odd. Thus $p - q - 1$ is always an even integer. Taking $z = 1$ and

$$w = \frac{p - 1 + q\sqrt{-19}}{2} = \frac{p - q - 1}{2} + \frac{q}{2}(1 + \sqrt{-19}) \in R, \text{ we see that}$$

$$\left(\frac{x}{y} \right) z - w = \frac{p + q\sqrt{-19}}{2} - \frac{p - 1 + q\sqrt{-19}}{2} = \frac{1}{2} \neq 0$$

$$\text{Hence } 0 < N \left\{ \left(\frac{x}{y} \right) z - w \right\} = \frac{1}{4} < 1.$$

Case III. $r = 3$.

Since $(p, q, r) = 1$, both p and q cannot be multiples of 3 simultaneously. Let $p = 3n + 1$ for some $n \in \mathbb{Z}$. Then for any $q \in \mathbb{Z}$, $p^2 + q^2 = 9n^2 + 6n + q^2 + 1 \Rightarrow$ if $3 \mid (p^2 + q^2)$ then $3 \mid (q^2 + 1)$. But $q = 3m$, $3m + 1$ or $3m + 2$ for some $m \in \mathbb{Z}$. For $q = 3m$, $3 \mid (q^2 + 1) \Rightarrow 3 \mid 1$! For $q = 3m + 1$, $3 \mid (q^2 + 1) \Rightarrow 3 \mid 2$! Now for $q = 3m + 2$, $3 \mid (q^2 + 1) \Rightarrow 3 \mid 5$! Hence $3 \nmid (p^2 + q^2)$. Similarly it can be shown that in case $p = 3m + 2$, $3 \nmid (p^2 + q^2)$. Further $p^2 + 19q^2 = (p^2 + q^2) + 18q^2 \Rightarrow 3 \nmid (p^2 + 19q^2)$. Taking $z = p - q\sqrt{-19}$ and $w = u \in R$, where $p^2 + 19q^2 = 3u + v$, $0 < v < 3$, (by our earlier discussion $v \neq 0$), we get

$$\left(\frac{x}{y} \right) z - w = \frac{(p + q\sqrt{-19})}{3}(p - q\sqrt{-19}) - u = \frac{v}{3} \neq 0.$$

$$\text{So } 0 < N \left\{ \left(\frac{x}{y} \right) z - w \right\} = \frac{v^2}{9} < 1, \text{ (as } v < 3).$$

Case IV. $r = 4$.

Since $(p, q, r) = 1$, both p and q cannot be even. Suppose p and

q are of opposite parity. Then $4 \nmid (p^2 - q^2)$ otherwise $2 \mid (p+q)$ or $2 \mid (p-q)$ which is absurd as both of $p+q$ and $p-q$ are odd. Also $p^2 + 19q^2 = (p^2 - q^2) + 20q^2 \Rightarrow 4 \nmid (p^2 + 19q^2)$. Take $z = p - q\sqrt{-19}$ and $w = u$ where $p^2 + 19q^2 = 4u + v$ with $0 < v < 4$. Then

$$\left(\frac{x}{y}\right)z - w = \frac{p^2 + 19q^2}{4} - u = \frac{v}{4} \neq 0.$$

$$\text{So } 0 < N \left\{ \left(\frac{x}{y}\right)z - w \right\} = \frac{v^2}{16} < 1.$$

Ultimately let p and q , both be odd. Then $p = 2m + 1$, $q = 2n + 1$ for some $m, n \in \mathbb{Z} \Rightarrow p^2 + 3q^2 = 4m(m+1) + 12n(n+1) + 4 \Rightarrow 4 \mid (p^2 + 3q^2)$ and $\frac{p^2 + 3q^2}{4} = m(m+1) + 3n(n+1) + 1$. Now $2 \mid m(m+1)$

and $2 \mid 3n(n+1) \Rightarrow 8 \nmid (p^2 + 3q^2)$, otherwise $2 \mid \left(\frac{p^2 + 3q^2}{4}\right) \Rightarrow 2 \mid 1$!

Further $p^2 + 19q^2 = (p^2 + 3q^2) + 16q^2 \Rightarrow 8 \nmid (p^2 + 19q^2)$. Then $p^2 + 19q^2 = 8l + j$ with $0 < j < 8$. Taking $z = \frac{p - q\sqrt{-19}}{2} = \frac{p+q}{2} - \frac{q}{2}(1 + \sqrt{-19}) \in R$ and $w = l$ we get $\left(\frac{x}{y}\right)z - w = \frac{p^2 + 19q^2}{8} - l = \frac{j}{8} \neq 0$.

$$\text{So } 0 < N \left\{ \left(\frac{x}{y}\right)z - w \right\} = \frac{j^2}{64} < 1.$$

Hence R is a PID. ■

Now it remains to prove that R is not a Euclidean Domain. Unless otherwise mentioned, in the following R is an integral domain with unity.

We shall denote by R_0 , the set of non-zero elements of R . Since R is an integral domain $R_0 \subseteq R$.

Definition 1. A subset P of R_0 is called a *Product Ideal* of R if $PR_0 \subseteq P$.

Note that $PR_0 = \{xy \mid x \in P, y \in R_0\}$ and R_0 is product ideal of R . If $P = \phi$, the empty set; then $PR_0 = \phi \subseteq \phi$ implies that ϕ is a product ideal of R . For any non-empty subset S of R , we put

$$S' = \{x \in S \mid y + xR \subseteq S \text{ for some } y \in R\}.$$

We denote $(S')'$ by S'' , $(S'')'$ by S''' and so on, in general we denote $(S^{(n-1)'})'$ by $S^{(n)}$ for any $n \in \mathbb{N}$.

Lemma 4. If S is a non-empty subset of R such that $S \subseteq T$ then $S' \subseteq T'$. Further if S is a product ideal of R then so is S' .

Proof. If $S' = \phi$, nothing remains to be proved. Suppose that $S' \neq \phi$ and let $x \in S'$. This implies that $x \in S$ and $\exists y \in R$ satisfying $y + xR \subseteq S$. But $S \subseteq T$ so $y + xR \subseteq T$. Also $x \in T$. Hence $x \in T'$. Thus $S' \subseteq T'$.

Now let S be a product ideal of R . If $S' = \phi$ then S' is a product

ideal of R . Suppose that $S' \neq \phi$ and let $x \in S'$. So $x \in S$ and $\exists y \in R$ such that $y + xR \subseteq S$. Let $z \in R_0$, since S is a product ideal $xz \in S \Rightarrow y + xzR \subseteq y + xR \subseteq S \Rightarrow xz \in S'$. Hence $S'R_0 \subseteq S'$. This prove that S' is a product ideal of R . ■

Lemma 5. *If R is a Euclidean Domain then there exists a sequence $\{P_n\}$ of product ideals with the following properties*

$$(i) R_0 = P_0 \supseteq P_1 \supseteq P_2 \supseteq P_3 \supseteq \dots \supseteq P_n \supseteq \dots,$$

$$(ii) \bigcap_{n=0}^{\infty} P_n = \phi,$$

$$(iii) P'_n \subseteq P_{n+1} \text{ for each } n \in \mathbb{N} \cap \{0\}, \text{ and}$$

$$(iv) \text{ for each } n \in \mathbb{N} \cup \{0\}, R_0^{(n)} \subseteq P_n.$$

Proof. Let δ be a Euclidean valuation on R . For each $n \in \mathbb{N} \cup \{0\}$, put $P_n = \{x \in R_0 \mid \delta(x) \geq n\}$. Note that if P_n is empty, P_n is a product ideal. In the other case let $x \in P_n$ and $z \in R_0$. Then $\delta(xz) \geq \delta(x) \Rightarrow \delta(xz) \geq n \Rightarrow xz \in P_n$. Hence $P_n R_0 \subseteq P_n$, i.e., P_n is a product ideal of R . Consider P_{n+1} . If $P_{n+1} = \phi$ then $P_n \supseteq P_{n+1}$. So let $P_{n+1} \neq \phi$ and let $x \in P_{n+1}$. This gives that $\delta(x) \geq n+1 \Rightarrow \delta(x) > n \Rightarrow x \in P_n$. Hence $P_n \supseteq P_{n+1}$, clearly $R_0 = P_0$; so we get (i).

Suppose that $\bigcap_{n=0}^{\infty} P_n \neq \phi$. Then there exists $x \in \bigcap_{n=0}^{\infty} P_n \Rightarrow x \in P_n \forall n \in \mathbb{N} \cup \{0\} \Rightarrow \delta(x) \geq n \forall n \in \mathbb{N} \cup \{0\}$. This is absurd as δ is a function from R_0 to $\mathbb{N} \cup \{0\}$. Hence $\bigcap_{n=0}^{\infty} P_n = \phi$. This proves (ii).

The case $P'_n = \phi$ is trivial. Assume that $P'_n \neq \phi$ and let $x \in P'_n$ then $\exists y \in R$ such that $y + xR \subseteq P_n$. As $x \neq 0$, there exist, by Euclidean Algorithm, $q, r \in R$ such that $y = xq + r$ with $r = 0$ or $\delta(r) < \delta(x)$. But $r = y + x(-q) \in y + xR \Rightarrow r \in P_n$. As a result of this $r \neq 0$ and $\delta(r) \geq n$. This forces that $\delta(x) > n \Rightarrow \delta(x) \geq n+1 \Rightarrow x \in P_{n+1}$. Hence $P'_n \subseteq P_{n+1}$. This proves (iii).

Finally $R_0 = P_0 \subseteq P_0 \Rightarrow R'_0 = P'_0 \subseteq P_1$ by (iii). So (iv) holds for $n=1$. Suppose that $R_0^{(n-1)} \subseteq P_{n-1}$. Then $R_0^{(n-1)'} \subseteq P_{n-1}'$ by Lemma 4. Using (iii) once again we obtain that $R_0^{(n)} \subseteq P_n$. Thus (iv) is proved. ■

Lemma 6. *If $R'_0 = R_0'' \neq \phi$ then R is not a Euclidean Domain.*

Proof: $R'_0 = R_0'' \Rightarrow R_0'' = (R'_0)' = (R_0'')' = R_0''' \Rightarrow R'_0 = R_0'''$. Proceeding in this manner we get that $R'_0 = R_0^{(n)}$ for all $n \in \mathbb{N}$. If R is a Euclidean Domain then by Lemma 5, we have $R'_0 = \bigcap_{n=1}^{\infty} R_0^{(n)} \subseteq \bigcap_{n=1}^{\infty} P_n \subseteq P_0 \forall n \in \mathbb{N} \Rightarrow \bigcap_{n=1}^{\infty} P_n \subseteq P_0$

$$\Leftrightarrow R'_0 \subseteq \bigcap_{n=0}^{\infty} P_n. \text{ Since by Lemma 5, } P_n \subseteq P_0 \forall n \in \mathbb{N} \Rightarrow \bigcap_{n=0}^{\infty} P_n \subseteq P_0$$

$\Rightarrow (\bigcap_{n=1}^{\infty} P_n) \cap P_0 = \bigcap_{n=1}^{\infty} P_n$. But by the same Lemma $\bigcap_{n=0}^{\infty} P_n = \phi$. Thus $R_0' = \phi$, which is against the hypothesis. Hence R is not a Euclidean Domain. ■

Lemma 7. $R_0' = \{x \in R \mid x \neq 0 \text{ and } x \text{ is non-unit}\}$.

Proof : As $R_0' \subseteq R_0$ and $0 \notin R_0'$ we get that $0 \notin R_0'$. Let x be a unit of R . There exists $y \in R$ such that $xy = 1$. For any $z \in R$, $z + x(-yz) = 0 \notin R_0' \Rightarrow x \notin R_0'$. Let $x \neq 0$ and x be a non-unit. Now $-1 + xy \neq 0 \forall y \in R$. $-1 + xy \in R_0' \forall y \in R \Rightarrow -1 + xR \subseteq R_0' \Rightarrow x \in R_0'$. Hence the Lemma follows. ■

Definition 2. As element x of R_0' is said to be a *side-divisor* of y in R if there exists $z \in R - R_0'$ such that $x \mid (y+z)$.

Definition 3. As element $x \in R_0'$ is called a *universal side-divisor* if it is a side-divisor of every element of R .

Lemma 8. $R_0'' = R_0' - U$, where U is the set of all universal side-divisors in R .

Proof : Let $R_0'' = \phi$ and let $x \in R_0'$ and $y \in R$. Since $x \notin R_0'' \Rightarrow -y + xR \not\subseteq R_0' \Rightarrow \exists t \in R$ such that $-y + xt \notin R_0'$. Now $x \mid y + (-y + xt) \Rightarrow x$ is a side-divisor of y . As y was chosen arbitrarily we get that x is a side-divisor of every element of R . Hence x is a universal side-divisor. This means that $R_0' \subseteq U$, i.e., $U = R_0'$. Consequently $R_0' - U = \phi$ and so $R_0'' = R_0' - U$.

Suppose now that $R_0'' \neq \phi$ and let $x \in R_0''$. Then $x \in R_0'$ and $y \in R$ such that $y + xR \in R_0'$. Suppose that $x \mid (y+z)$ for some $z \notin R_0'$. This means that $y+z = kx$ for some k in R and z is either 0 or a unit (Lemma 7). Now $y + x(-k) = -z$ and $y + xR \subseteq R_0' \Rightarrow -z \in R_0' \Rightarrow -z \neq 0$ and $-z$ is a non-unit $\Rightarrow z \neq 0$ and z is a non-unit. This is against the hypothesis $x \mid (y+z)$ for some $z \in R - R_0'$. Hence $x \notin U$. This implies $x \in R_0' - U$ and so $R_0'' \subseteq R_0' - U$. Conversely let $x \in R_0' - U$. If $x \notin R_0''$, then for each y in R , $y + xR \not\subseteq R_0'$. In other words for each y in R $\exists w$ in R such that $y + xw \notin R_0'$. So either $y + xw = 0$ or $y + xw$ is a unit. Thus forces $y + xw = -z$ where either $z = 0$ or z is a unit. Thus $y + z = -xw$ with $z \in R - R_0' \Rightarrow x \mid (y+z)$ in R . Hence x is a side-divisor of y . Since this holds for all y in R , x is a universal side-divisor. But $x \notin U$, so we get a contradiction. Hence $x \in R_0' - U \Rightarrow x \in R_0''$. Consequently $R_0' - U \subseteq R_0''$ and then it follows at once that $R_0'' = R_0' - U$. ■

Theorem 2. $R = \left\{ a + \frac{b}{2} (1 + \sqrt{-19}) \mid a, b \in \mathbf{Z} \right\}$ is not a Euclidean Domain.

Proof: In view of Lemma 6, it is sufficient to prove that

$R_0'' = R_0' \neq \phi$. For this purpose we show that the set U is a null set. Then in view of Lemma 8, $R_0'' = R_0'$ and as R_0 has non-zero non-units, this will in turn yield that $R_0'' = R_0' \neq \phi$.

Consider a side-divisor of 2. If $x \in R_0'$ is such that $x \mid (2+z)$ for some z in $R - R_0'$ then either $z=0$ or 1 or -1 . In case $z=-1$ we get $x \mid 1 \Rightarrow x$ is a unit $\Rightarrow x \notin R_0'$. This is absurd!

In case $z=0$ we have $x \mid 2$ and in case $z=1$ we have $x \mid 3$.

We claim that 2 and 3 are irreducible elements of R . Suppose

$$2 = \left\{ a + \frac{b}{2} (1 + \sqrt{-19}) \right\} \left\{ c + \frac{d}{2} (1 + \sqrt{-19}) \right\} \text{ with } a, b, c, d \in \mathbb{Z}.$$

$$\text{Then } N(2) = N \left\{ a + \frac{b}{2} (1 + \sqrt{-19}) \right\} N \left\{ c + \frac{d}{2} (1 + \sqrt{-19}) \right\} \Rightarrow$$

$$4 = \left[\left(a + \frac{b}{2} \right)^2 + \frac{19b^2}{4} \right] \left[\left(c + \frac{d}{2} \right)^2 + \frac{19d^2}{4} \right]. \text{ If } b \neq 0 \text{ then}$$

$$\frac{19b^2}{4} \geq \frac{19}{4} > 4 \Rightarrow \left(a + \frac{b}{2} \right)^2 + \frac{19b^2}{4} > 4. \text{ This is impossible. Hence}$$

$b=0$. Similarly it can be proved that $d=0$. So we obtain $2=ac$. This in turn implies that either $a=2, c=1$, or $a=-2, c=-1$, or $a=1, c=2$; or $a=-1, c=-2$. In each case one of the two factors of 2 is a unit, hence 2 is an irreducible element of R .

$$\text{Again } 3 = \left\{ a + \frac{b}{2} (1 + \sqrt{-19}) \right\} \left\{ c + \frac{d}{2} (1 + \sqrt{-19}) \right\} \text{ with}$$

$$a, b, c, d \in \mathbb{Z} \Rightarrow 9 = \left[\left(a + \frac{b}{2} \right)^2 + \frac{19b^2}{4} \right] \left[\left(c + \frac{d}{2} \right)^2 + \frac{19d^2}{4} \right]. \text{ If}$$

$$b \neq 0 \text{ then } \frac{19b^2}{4} \geq \frac{19}{4} > 3 \Rightarrow \left(a + \frac{b}{2} \right)^2 + \frac{19b^2}{4} = 9. \text{ (The only}$$

factor of 9 in $\mathbb{Z}$, which is greater than 3, is 9 itself). Then

$$\left(c + \frac{d}{2} \right)^2 + \frac{19d^2}{4} = 1. \text{ This implies that } d=0 \text{ and } c=\pm 1. \text{ In other}$$

$$\text{words } c + \frac{d}{2} (1 + \sqrt{-19}) \text{ is a unit. If } b=0 \text{ then } a^2 \left[\left(c + \frac{d}{2} \right)^2 + \frac{19d^2}{4} \right] = 9 \Rightarrow a^2 = 1 \text{ or } 9. \text{ If } a^2 = 9 \text{ then again we get}$$

$$\left(c + \frac{d}{2} \right)^2 + \frac{19d^2}{4} = 1 \Rightarrow d=0 \Rightarrow c=\pm 1 \text{ and } c + \frac{d}{2} (1 + \sqrt{-19}) \text{ is a unit. If } a^2 = 1 \text{ then}$$

$$a=\pm 1 \text{ and } b=0 \text{ together imply that } a + \frac{b}{2} (1 + \sqrt{-19}) \text{ is a unit.}$$

Hence 3 is a irreducible element of R .

Now as a side-divisor of 2 is in R_0' , it must be a non-unit. Since for $z=0$ or $z=1$ it either divides 2 or 3 in R , we get that the only side-divisors of 2 are ± 2 and ± 3 .

Now a side-divisor of $\frac{1+\sqrt{-19}}{2} \in R$ must be a non-unit divisor of $\frac{1+\sqrt{-19}}{2}$ or $\frac{1+\sqrt{-19}}{2} + 1 = \frac{3+\sqrt{-19}}{2}$ or $\frac{1+\sqrt{-19}}{2} - 1 = \frac{-1+\sqrt{-19}}{2}$. The norms of these are 5, 7 and 5 respectively.

Further we know that if $x(\neq 0)$ divides y then $N(x) \mid N(y)$ and norms of 2 and 3 are 4 and 9 respectively. So no side-divisor of 2 can be a side-divisor of $\frac{1+\sqrt{-19}}{2}$.

Consequently if R_0 has a universal side-divisor, say u . Then u must be a side-divisor of 2 as well as a side-divisor of $\frac{1+\sqrt{-19}}{2}$.

This is a contradiction to what we have proved above. Hence R_0' has no universal side-divisor. Then by our statement in the beginning of the proof, R is not a Euclidean Domain. ■

3. FUNDAMENTAL THEOREM OF ALGEBRA

In this part of the appendix we give a proof of "Fundamental Theorem of Algebra". We assume, of course, that the readers are acquainted with the elementary results of Real Analysis. First of all we state and prove Intermediate-Value Theorem, usually attributed to Weierstrass.

Theorem 1 (Intermediate-Value Theorem). $f(x)$ is a real-valued function, continuous on the closed interval $[a, b]$ with $f(a) < 0$ and $f(b) > 0$. There exists at least one $c \in]a, b[$ such that $f(c) = 0$.

Proof : Suppose that the theorem is false. Determine real numbers a_n and b_n ($n=1, 2, \dots$) such that

$$(i) \ a_1 = a, \ b_1 = b$$

$$(ii) \ a_{n+1} = a_n, \ b_{n+1} = \frac{1}{2}(a_n + b_n); \text{ in case } f[\frac{1}{2}(a_n + b_n)] \text{ is positive.}$$

$$(iii) \ b_{n+1} = b_n, \ a_{n+1} = \frac{1}{2}(a_n + b_n); \text{ in case } f[\frac{1}{2}(a_n + b_n)] \text{ is negative.}$$

By our choice, it is evident that $a_1 \leq a_2 \leq a_3 \leq \dots$ and $b_1 \geq b_2 \geq b_3 \geq \dots$. Further $a_i \leq b \ \forall i \in \mathbb{N}$ and $b_j \geq a \ \forall j \in \mathbb{N}$. Thus $\{a_n\}$ is a monotonic increasing sequence which is bounded above and $\{b_n\}$ is a monotonic decreasing sequence which is bounded below. Consequently both the sequences are convergent. Since $b_n - a_n$

$$= \frac{1}{2^{n-1}} (b-a) \rightarrow 0 \text{ as } n \rightarrow \infty, \text{ we get that } \lim_{n \rightarrow \infty} a_n = \lim_{n \rightarrow \infty} b_n = c, \text{ say.}$$

As $\{a_n\}$ is monotonically increasing and $\{b_n\}$ is monotonically decreasing $a_n \leq c \leq b_n \ \forall n \in \mathbb{N}$. Hence $c \in [a, b]$. Also note that $f(a_n) < 0$

and $f(b_n) > 0 \forall n \in \mathbf{N}$. Because of our assumption $f(c) \neq 0$.

Let $f(c) > 0$. As $f(x)$ is continuous at $x=c$, there exists $\delta \in \mathbf{R}$ ($\delta > 0$), such that $f(x) > 0 \forall x$ in the open interval $]c-\delta, c+\delta[$.

Now $\lim_{n \rightarrow \infty} a_n = c \Rightarrow \exists n_1 \in \mathbf{N}$ such that $|a_n - c| < \delta$ for all $n \geq n_1$.

This implies that $a_n \in]c-\delta, c+\delta[\forall n \geq n_1$. Similarly $\lim_{n \rightarrow \infty} b_n = c$

$\Rightarrow \exists n_2 \in \mathbf{N}$ such that $b_n \in]c-\delta, c+\delta[\forall n \geq n_2$. Let $n_3 = \max(n_1, n_2)$. Then both a_n and b_n belong to $]c-\delta, c+\delta[$ for all $n \geq n_3$. As $f(a_n) < 0$ and $f(b_n) > 0$ for all $n \geq n_3$, we get a contradiction.

Similarly $f(c) < 0$ leads to a contradiction. ■

Theorem 2. Every polynomial $f(x) \in \mathbf{R}[x]$ of odd degree, has at least one root in $\mathbf{R}$.

Proof: We may, without loss of generality, assume that $f(x)$ is a monic polynomial.

Let $f(x) = x^n + \alpha_1 x^{n-1} + \alpha_2 x^{n-2} + \dots + \alpha_n$ with α_i 's $\in \mathbf{R}$ and n , an odd integer.

Choose a real number c such that $c > 1$ and $c > |\alpha_1| + |\alpha_2| + \dots + |\alpha_n|$.

Now $f(c) = c^n + \alpha_1 c^{n-1} + \alpha_2 c^{n-2} + \dots + \alpha_n$

$$\begin{aligned} &= c^n \left\{ 1 + \frac{\alpha_1}{c} + \frac{\alpha_2}{c^2} + \dots + \frac{\alpha_n}{c^n} \right\} \\ &\geq c^n \left\{ 1 - \left| \frac{\alpha_1}{c} + \frac{\alpha_2}{c^2} + \dots + \frac{\alpha_n}{c^n} \right| \right\} \\ &\geq c^n \left\{ 1 - \left(\frac{|\alpha_1|}{c} + \frac{|\alpha_2|}{c^2} + \dots + \frac{|\alpha_n|}{c^n} \right) \right\} \end{aligned}$$

as $\left| \frac{\alpha_1}{c} + \frac{\alpha_2}{c^2} + \dots + \frac{\alpha_n}{c^n} \right| \leq \frac{|\alpha_1|}{c} + \frac{|\alpha_2|}{c^2} + \dots + \frac{|\alpha_n|}{c^n}$

This together with the fact that $c > 1$ implies that

$$f(c) \geq c^n \left\{ 1 - \left(\frac{|\alpha_1|}{c} + \frac{|\alpha_2|}{c} + \dots + \frac{|\alpha_n|}{c} \right) \right\}.$$

As $c > \sum_{i=1}^n |\alpha_i|$ we get that $f(c) > 0$.

Again $f(-c) = -c^n + \alpha_1 c^{n-1} - \alpha_2 c^{n-2} + \dots + \alpha_n$, as n is odd

$$= -(c^n - \beta_1 c^{n-1} + \beta_2 c^{n-2} + \dots + \beta_n)$$

where $\beta_i = (-1)^i \alpha_i \forall i = 1, 2, \dots, n$.

As $|\beta_i| = |\alpha_i| \forall i = 1, 2, \dots$, we see that $c > \sum_{i=1}^n |\beta_i|$.

Using the arguments given earlier, it can be proved that

$$-f(-c) = c^n + \beta_1 c^{n-1} - \dots + \beta_n > 0 \text{ or } f(-c) < 0.$$

Clearly $f(x)$ is a real-valued continuous function of x . Using

Theorem 1 we get that there exists at least one real number $d \in]-c, c[$ such that $f(d) = 0$. ■

Lemma. 3. Every polynomial in $\mathbf{C}[x]$ of degree 2, has all its roots in $\mathbf{C}$.

Proof : Let $ax^2 + bx + c \in \mathbf{C}[x]$ with $a \neq 0$. We can write

$$\begin{aligned} ax^2 + bx + c &= \frac{1}{a} \left[a^2x^2 + abx + ac \right] \\ &= \frac{1}{a} \left[\left(ax + \frac{b}{2} \right)^2 - \frac{b^2 - 4ac}{4} \right] \\ &= \frac{1}{a} \left(ax + \frac{b}{2} + \frac{\sqrt{b^2 - 4ac}}{2} \right) \left(ax + \frac{b}{2} - \frac{\sqrt{b^2 - 4ac}}{2} \right) \end{aligned}$$

Thus the roots of $ax^2 + bx + c$ are $\frac{-b + \sqrt{b^2 - 4ac}}{2a}$ and $\frac{-b - \sqrt{b^2 - 4ac}}{2a}$ in any extension of $\mathbf{C}$. We show that $\sqrt{b^2 - 4ac} \in \mathbf{C}$.

Let $b^2 - 4ac = \alpha + i\beta$ where $\alpha, \beta \in \mathbf{R}$. Consider real numbers $\frac{1}{2}(\alpha + \sqrt{\alpha^2 + \beta^2})$ and $\frac{1}{2}(\sqrt{\alpha^2 + \beta^2} - \alpha)$. As both of them are positive, we can find real numbers γ and δ such that $\gamma^2 = \frac{1}{2}(\alpha + \sqrt{\alpha^2 + \beta^2})$ and $\delta^2 = \frac{1}{2}(\sqrt{\alpha^2 + \beta^2} - \alpha)$. Then $\gamma^2 + \delta^2 = \sqrt{\alpha^2 + \beta^2}$ and $\gamma^2 - \delta^2 = \alpha$. So $\alpha^2 + \beta^2 = (\gamma^2 + \delta^2)^2 \Rightarrow \beta^2 = -\alpha^2 + (\gamma^2 + \delta^2)^2 = -(\gamma^2 - \delta^2)^2 + (\gamma^2 + \delta^2)^2 \Rightarrow \beta = \pm 2\gamma\delta$. Hence $b^2 - 4ac = \gamma^2 - \delta^2 \pm 2\gamma\delta i = (\gamma \pm \delta i)^2 \Rightarrow \sqrt{b^2 - 4ac} = \gamma \pm \delta i \in \mathbf{C}$. Consequently both the roots of $ax^2 + bx + c$ are in $\mathbf{C}$. ■

For any polynomial $f(x) = \sum_{i=0}^n \alpha_i x^i \in \mathbf{C}[x]$ we denote by $\bar{f}(x)$, the polynomial $\sum_{i=0}^n \bar{\alpha}_i x^i$, where for each $i \in \{0, 1, 2, \dots, n\}$ $\bar{\alpha}_i$ is complex conjugate of α_i . It can be checked that $f(x) \bar{f}(x) \in \mathbf{R}[x]$.

Lemma 4. For any non-constant polynomial $f(x) \in \mathbf{C}[x]$, the polynomial $g(x) = f(x) \bar{f}(x) \in \mathbf{R}[x]$ has a root in $\mathbf{C}$ if and only if $f(x)$ has a root in $\mathbf{C}$.

Proof : Let $c \in \mathbf{C}$ be a root of $f(x)$ then $f(c) = 0$ implies that $g(c) = 0$. Hence c is a root of $g(x)$.

Conversely let $c \in \mathbf{C}$ be a root of $g(x)$. Then $g(c) = 0 \Rightarrow f(c) \bar{f}(c) = 0$. In case $f(c) = 0$, nothing remains to be proved. In case $\bar{f}(c) = 0$, let

$f(x) = \sum_{i=1}^n \alpha_i x^i$ then $\bar{f}(x) = \sum_{i=1}^n \bar{\alpha}_i x^i$. Now $\bar{f}(c) = 0 \Rightarrow \sum_{i=1}^n \bar{\alpha}_i c^i = 0$
 $\Rightarrow \sum_{i=0}^n \alpha_i \bar{c}^i = 0 \Rightarrow \bar{c}$ is a root of $f(x)$. As $\bar{c} \in \mathbf{C}$ we get that $f(x)$ has a root $\bar{c}$ in $\mathbf{C}$. ■

Now we are ready for the famous theorem.

Theorem 5. (Fundamental Theorem of Algebra). *The field of complex numbers is algebraically closed.*

Proof: In view of Lemma 4 it is sufficient to prove that any non-constant polynomial $R[x]$ has at least one root in $\mathbf{C}$. Let $f(x) \in R[x]$ be of degree $d > 1$. We can write $d = 2^m q$ for some odd integer q and $m \in \mathbf{N} \cup \{0\}$. We shall apply induction on m .

For $m=0$, $f(x) \in R[x]$ is of odd degree. Hence it has at least one root in $\mathbf{R}$ (Theorem 2). As $\mathbf{R} \subset \mathbf{C}$, we get the result.

Now let $m \geq 1$ and let the theorem be true for all polynomials in $R[x]$ of degree $2^r s$ with $r \in \mathbf{N} \cup \{0\}$, $r < m$ and s , an odd integer.

Consider $f(x)$. Let K be the splitting field of $f(x)$ over $\mathbf{R}$ and let $x_1, x_2, \dots, x_d$ be all the roots of $f(x)$. Then $f(x) = \prod_{i=1}^d (x - x_i)$. Let c

be an arbitrary element of $\mathbf{R}$. Put $y_{ij} = x_i + x_j + cx_i x_j$, ($1 \leq i \leq j \leq d$). Clearly $y_{ii} \in K$. Since each x_i can be chosen in d ways and for $x_i \neq x_j$, the two elements $x_i + x_j + cx_i x_j$, $x_j + x_i + cx_j x_i$ are equal, we get the number of y_{ij} 's (with $x_i \neq x_j$) is equal to $\frac{d(d-1)}{2}$. The

number of y_{ij} 's of type $x_i + x_i + cx_i^2$ is d .

Hence the total number of elements y_{ij} ($1 \leq i \leq j \leq d$) is equal to $\frac{d(d-1)}{2} + d$, i.e., $\frac{d(d+1)}{2} = 2^{m-1}q(d+1)$. Since d is even integer,

$d+1$ is odd and consequently $q(d+1)$ is an odd integer.

The polynomial $g(x) = \prod_{i,j} (x - y_{ij})$ is of degree $2^{m-1}q(d+1)$. The

coefficients of $g(x)$ are elementary symmetric functions of y_{ij} 's over $\mathbf{R}$. Each y_{ij} is a polynomial in x_i 's over $\mathbf{R}$. Thus each coefficient in

$g(x)$ is a symmetric polynomial of x_i 's over $\mathbf{R}$. Consequently coefficients of $g(x)$ are polynomials of elementary symmetric func-

tions of x_i 's over $\mathbf{R}$. Let $\sigma_1 = \sum x_i$, $\sigma_2 = \sum_{i>j} x_i x_j$, $\sigma_3 = \sum_{i>j>k} x_i x_j x_k$,

$\dots$, $\sigma_d = x_1 x_2 \dots x_d$ be the elementary symmetric functions of x_i 's.

Since $f(x) \in R[x]$, all the σ_i 's $\in \mathbf{R}$. Consequently $g(x) \in R[x]$. By induction hypothesis there is a complex number say z_c , for which $g(z_c) = 0$. Hence there exist integers $i(c), j(c) \in \{1, 2, \dots, d\}$ such that

$z_c = x_{i(c)} + x_{j(c)} + cx_{i(c)} x_{j(c)}$.

We know that the set of ordered pairs (i, j) with $1 \leq i, j \leq d$, is finite while $\mathbf{R}$ is an infinite set. So there exist two distinct real num-

bers c_1 and c_2 with $i(c_1) = i(c_2)$ and $s = j(c_1) = j(c_2)$. For convenience sake we put $r = i(c_1) = i(c_2)$ and $s = j(c_1) = j(c_2)$. Thus $z_{c_1} = x_r + x_s$

$c_1x_r x_s \in \mathbf{C}$ and $z_{c_2} = x_r + x_s + c_2 x_r x_s \in \mathbf{C}$. This gives that $(c_1 - c_2)x_r x_s \in \mathbf{C} \Rightarrow x_r x_s \in \mathbf{C}$ as $c_1 - c_2 \neq 0 \in \mathbf{R} \subset \mathbf{C}$. This in turn yields that $x_r + x_s = z_{c_1} - c_1 x_r x_s \in \mathbf{C}$. Hence both $x_r + x_s$ and $x_r x_s \in \mathbf{C}$. But x_r, x_s are the roots of polynomial $x^2 - (x_r + x_s)x + x_r x_s \in \mathbf{C}[x]$. By Lemma 3 its roots are in $\mathbf{C}$. In other words $x_r, x_s \in \mathbf{C}$. Hence at least one of the roots of $f(x)$ is in $\mathbf{C}$. ■

Note. Rigorous proofs of Theorem 5 were given by d' Alembert and Gauss for first time. That is why, this theorem is also known as d' Alembert and Gauss, Theorem.

4. TRANSCENDENCE OF e

In this part we prove that e is a transcendental number, that is, e does not satisfy any non-zero polynomial in $\mathbf{Z}[x]$.

Firstly we prove a Lemma.

Lemma 1. Let $h(x) = \frac{x^n g(x)}{n!}$ where $g(x) \in \mathbf{Z}[x]$. Then $h^{(j)}(0)$, the j^{th} derivative of $h(x)$ evaluated at $x=0$ is an integer for $j=0, 1, 2, \dots$. Moreover, with the possible exception of the case $j=n$, the integers $h^{(j)}(0)$ are divisible by $n+1$. If x divides $g(x)$ then $h^{(n)}(0)$ is also divisible by $n+1$.

Proof: Let c_j be the coefficient of x^j in $x^n g(x)$. Note that $h^{(j)}(0) = \frac{1}{n!} [c_j (j!)]$. As $g(x) \in \mathbf{Z}[x]$, $c_j \in \mathbf{Z}$. Clearly there is no term x^j in $h(x)$ for $j < n$. So $h^{(j)}(0) = 0 \forall j = 1, 2, \dots, n-1$. For $j=0$, $h^{(0)}(x) = h(x)$ (by definition), so we have $h^{(0)}(0) = 0$. Hence $h^{(j)}(0) \in \mathbf{Z}$ and is divisible by $n+1$ for all $j=0, 1, \dots, n-1$. If $j \geq n$ then $n!$ divides $j!$ and so $h^{(j)}(0)$ is an integer. In case $j > n$, $\frac{j!}{n!} = j(j-1)\dots(n+1) \Rightarrow (n+1) \mid \frac{j!}{n!} \Rightarrow (n+1) \mid h^{(j)}(0)$. Hence the first part of the Lemma follows.

For $j=n$, $h^{(n)}(0) = c_n$. If x divides $g(x)$ then $g(x) = xq(x)$ for some $q(x) \in \mathbf{Z}[x] \Rightarrow h(x) = \frac{x^{n+1}q(x)}{n!} \Rightarrow c_n = 0 \Rightarrow (n+1) \mid c_n$. ■

Theorem 2. e is a transcendental number.

Proof: (Niven). Let $a_m e^m + a_{m-1} e^{m-1} + \dots + a_1 e + a_0 = 0 \dots (A)$ for $a_i \in \mathbf{Z}$ with at least one of a_i 's, a non-zero integer. We can assume, without loss of generality, that $a_0 \neq 0$.

Put $f(x) = \frac{x^{p-1}(x-1)^p(x-2)^p \dots (x-m)^p}{(p-1)!}$ where p is an odd prime to be specified later.

Let $F(x) = f(x) + f'(x) + f^{(2)}(x) + \dots + f^{(mp+p-1)}(x)$
 For $0 < x < m$, we have

$$|f(x)| < \frac{m^{p-1} m^p m^p \dots m^p}{(p-1)!} = \frac{m^{mp+p-1}}{(p-1)!} \quad \dots(1)$$

$$\begin{aligned} \text{Now } \frac{d}{dx} [e^{-x} F(x)] &= e^{-x} F'(x) - e^{-x} F(x) = e^{-x} [F'(x) - F(x)] \\ &= e^{-x} [f'(x) + f^{(2)}(x) + \dots + f^{(mp+p)}(x) - f(x) \\ &\quad - f'(x) - f^{(2)}(x) - \dots - f^{(mp+p-1)}(x)]. \end{aligned}$$

$$\text{So } \frac{d}{dx} [e^{-x} F(x)] = -e^{-x} f(x). \quad \dots(2)$$

[As $f(x)$ is of degree $mp+p-1$, $f^{mp+p}(x) = 0$].

Again

$$\begin{aligned} a_j \int_0^j e^{-x} f(x) dx &= a_j \left[-e^{-x} F(x) \right]_0^j = a_j \left[-e^{-j} F(j) + F(0) \right] \\ &= a_j F(0) - a_j e^{-j} F(j). \end{aligned} \quad \dots(3)$$

Multiply (3) by e^j and sum over $j = 0, 1, 2, \dots$.

$$\begin{aligned} \text{Then } \sum_{j=0}^m a_j e^j \int_0^j e^{-x} f(x) dx &= F(0) \sum_{j=0}^m a_j e^j - \sum_{j=0}^m a_j e^j F(j) \\ &= - \sum_{j=0}^m a_j F(j), \text{ by (A)} \\ &= - \sum_{j=0}^m a_j \left[\sum_{i=0}^{mp+p-1} f^{(i)}(j) \right] \\ &= - \sum_{j=0}^m \sum_{i=0}^{mp+p-1} a_j f^{(i)}(j) \quad \dots(4) \end{aligned}$$

Apply Lemma 1 in succession to $f(x), f(x+1), f(x+2), \dots, f(x+m)$, with h replaced by f and n replaced by $p-1$ in each case. Thus $f^{(i)}(j)$ is an integer for all values of i and j in the sum on right side of (4). Again, by Lemma 1, $f^{(i)}(j)$ is divisible by p except for the case when $j=0$ and $i=p-1$.

Now coefficient of x^{p-1} in $f(x)$ is

$$\frac{(-1)^p (-2)^p \dots (-m)^p}{(p-1)!}.$$

$$\begin{aligned} \text{Hence } f^{(p-1)}(0) &= \frac{(-1)^p (-2)^p \dots (-m)^p [(p-1)!]}{(p-1)!} \\ &= (-1)^p (-2)^p \dots (-m)^p. \end{aligned}$$

Thus if we choose $p > m$, then $f^{(p-1)}(0)$ is not divisible by p . Further if p is chosen to be greater than $|a_0|$ then right side of (4) consists of a sum of multiples of p with one exception, namely $-a_0 f^{(p-1)}(0)$. Therefore the sum on the right side of (4) is a non-zero integer.

But the left side of (4) satisfies the inequality

$$\begin{aligned} \left| \sum_{j=0}^m a_j e^j \int_0^j f(x) e^{-x} dx \right| &\leq \sum_{j=0}^m \left| a_j e^j \int_0^j e^{-x} f(x) dx \right| \\ &\leq \left[\sum_{j=0}^m |a_j| j e^j \right] \frac{m^{m+p-1}}{(p-1)!} \\ &\leq \left[\sum_{j=0}^m |a_j| \right] e^m m \frac{m^{m+p-1}}{(p-1)!} \\ &\leq \left[\sum_{j=0}^m |a_j| \right] e^m m^{m+1} \frac{(m^{m+1})^{p-1}}{(p-1)!} \\ &\leq 1 \end{aligned}$$

as p can be chosen as large as we like.

This is a contradiction. ■

BIBLIOGRAPHY

- Adamson, I. T., *Elementary Rings and Modules*. Edinburgh: Oliver & Boyd, 1972.
- , *Introduction to Field Theory*. London: Oliver & Boyd, 1964.
- Albert, A. A., *Fundamental Concepts of Higher Algebra*. Chicago: University of Chicago Press, 1956.
- Ames, D. B., *An Introduction to Abstract Algebra*. Scranton, Pennsylvania: International Text Book Company, 1969.
- , *Fundamentals of Linear Algebra*. Scranton, Pennsylvania: International Text Book Company, 1970.
- Artin, E., *Galois Theory*. Notre Dame, Indiana: University of Notre Dame Press, 1955.
- , C. Nesbitt and R. Thrall, *Rings with Minimum Condition*. Ann. Arbor, Michigan: University of Michigan Press, 1944.
- Atiyah, M. F. and I. G. Macdonald, *Introduction to Commutative Algebra*. Reading, Massachusetts: Addison-Wesley, 1969.
- Auslander, M. and D. A. Buchsbaum, *Groups, Rings, Modules*. New York: Harper & Row, 1974.
- Barnes, W. E., *Introduction to Abstract Algebra*. Boston: Heath, 1963.
- Barshay, J., *Topics in Ring Theory*. New York: W. A. Benjamin, Inc., 1969.
- Baumslag, B. and B. Chandler, *Theory and Problems of Group Theory*. New York: McGraw-Hill Book Company, 1968.
- Behrens, E. A., *Ring Theory*. New York: Academic Press, 1972.
- Birkhoff, G. and S. MacLane, *A Survey of Modern Algebra*. New York: Macmillan, 1965.
- , *Algebra*. New York: Macmillan, 1979.
- Bourbaki, N., *Commutative Algebra*. Reading, Massachusetts: Addison-Wesley, 1972.
- Burnside, W., *Theory of Groups of Finite Order*. New York: Dover, 1955.
- Burton, D. M., *A First Course in Rings and Ideals*. Reading, Massachusetts: Addison-Wesley, 1968.
- , *Abstract and Linear Algebra*. Addison-Wesley 1972.
- Charmichael, R. D., *Introduction to the Theory of Groups of Finite Order*. New York: Dover, 1956.
- Chevalley, C., *Fundamental Concepts of Algebra*. New York: Academic Press, 1956.
- Cohn, P. M., *Algebra*, Vol. I. New York: John Wiley & Sons, 1974.
- , *Algebra*, Vol. II. New York: John Wiley & Sons, 1977.
- Dean, R. A., *Elements of Abstract Algebra*. New York: Wiley, 1966.
- Dixon, J. D., *Problems in Group Theory*. Waltham, Massachusetts: Blaisdell, 1967.
- Dubreil, P. and M. L. Dubreil Jacotin, *Lectures on Modern Algebra*. New York: Hafner Publishing Company, 1967.
- Durbin, J. R., *Modern Algebra, An Introduction*. New York: John Wiley, 1979.

- Finkbeiner II, D. T., *Introduction to Matrices and Linear Transformations*. Tokyo: Toppan Company Ltd., 1965.
- Fraleigh, J. B., *A First Course in Abstract Algebra*. Reading, Massachusetts: Addison-Wesley, 1968.
- Fuchs, L., *Abelian Groups*. Oxford: Pergamon Press, 1960.
- , *Infinite Abelian Groups*, Vol. I. New York: Academic Press, 1970.
- , *Infinite Abelian Groups*, Vol. II. New York: Academic Press, 1973.
- Gaal, L., *Galois Theory with Examples*. New York: Chelsea, 1973.
- Goldhaber, J. K. and G. Ehrlich, *Algebra*. London: Macmillan, 1970.
- Goldstein, L. R., *Abstract Algebra, a First Course*. Englewood Cliffs, New Jersey: Prentice-Hall, 1973.
- Gorenstein, D., *Finite Groups*. New York: Harper and Row, 1968.
- Gray, M., *A Radical Approach to Algebra*. Reading, Massachusetts: Addison-Wesley, 1970.
- Green, J. A., *Sets and Groups*. London: Routledge and Kegan Paul, 1965.
- Greub, W. H., *Linear Algebra*. Berlin: Springer-Verlag, 1967.
- Griffith, P. A., *Infinite Abelian Group Theory*. Chicago: The University of Chicago Press, 1970.
- Hall Jr., M., *The Theory of Groups*. New York: Macmillan, 1959.
- Halmos, P., *Finite-dimensional Vector Spaces*. Princeton, New Jersey: Von Nostrand, 1958.
- , *Naive Set Theory*. Princeton, New Jersey: Von Nostrand, 1960.
- Herstein, I. N., *Topics in Algebra*. Delhi: Vikas, 1971.
- Hilton, P. J. and Y-C. Wu, *A Course in Modern Algebra*. New York: John Wiley, 1974.
- Hoffman, K. and R. Kunze, *Linear Algebra*. Englewood Cliffs; New Jersey: Prentice-Hall, 1961.
- Huneford, T. W., *Algebra*. New York: Holt, Rinehart and Winston, 1974.
- Huppert, B., *Endliche Gruppen I*, Berlin: Springer-Verlag, 1967.
- Hu, S. T., *Elements of Modern Algebra*. San Francisco; California: Holden Day Inc., 1965.
- Jacobson, N., *Basic Algebra*. Vol. I. San Francisco: W. H. Freeman, 1974.
- , *Lectures in Abstract Algebra*, Vol. I. Princeton, New Jersey: Von Nostrand, 1951.
- , *Lectures in Abstract Algebra*, Vol. II. Von Nostrand, 1953.
- , *Lectures of Abstract Algebra*, Vol. III. Von Nostrand, 1964.
- , *Structure of Rings*. Providence. American Mathematical Society, 1964.
- Kaplansky, I., *Commutative Rings*. Boston, Massachusetts: Allyn & Bacon, 1970.
- , *Fields and Rings*. Chicago: The University of Chicago Press, 1969.
- Kargapolov, M. I. and Ju I. Merzljakov, *Fundamentals of the Theory of Groups*. New York: Springer-Verlag, 1979.
- Kochendorfer, R., *Group Theory*. London: McGraw-Hill, 1970.
- Kuczowski, J. E. and J. L. Gerstins, *Abstract Algebra, a First Look*. New York: Marcel Dekker, 1977.
- Kurosh, A. G., *The Theory of Groups*, (Two volumes). New York: Chelsea, 1960.
- Lambek, J., *Lectures on Rings and Modules*. Waltham, Massachusetts: Blaisdell, 1966.
- Lang, S., *Algebra*. Reading, Massachusetts: Addison-Wesley, 1965.

- , *Linear Algebra*. Reading, Massachusetts: Addison-Wesley, 1968.
- Ledermann, W., *An Introduction to the Theory of Finite Groups*. London: Oliver & Boyd., 1964.
- , *Introduction to Group Theory*. London: Longman, 1973.
- Lewis, D. J., *Introduction to Algebra*. New York: Harper & Row, 1965.
- Macdonald, I. D., *The Theory of Groups*. Oxford: Oxford University Press, 1968.
- Macduffee, C. C., *An Introduction to Abstract Algebra*. New York: Dover, 1964.
- Marcus, M., *Introduction to Modern Algebra*. New York: Marcel Dekker, 1978.
- McCarthy, P., *Algebraic Extension of Fields*. Waltham, Massachusetts: Blaisdell, 1966.
- McCoy, N. H., *The Theory of Rings*. New York: Macmillan, 1964.
- , *Introduction to Modern Algebra*. Boston: Allyn & Bacon, 1975.
- , *Fundamentals of Abstracts Algebra*. Allyn & Bacon, 1972.
- , and T. R. Berger, *Algebra, Groups, Rings and other Topics*. Allyn & Bacon, 1977.
- Moore, J. T., *Elements of Abstract Algebra*. New York: Macmillan, 1967.
- , *Introduction to Abstract Algebra*. New York: Academic Press, 1975.
- Mostow, G. D., J. H. Sampson and J. Meyer, *Fundamental Structures of Algebra*. New York: McGraw-Hill, 1963.
- Nagata, M., *Field Theory*. New York: Marcel Dekker, 1977.
- Niven, I., *Irrational Numbers*. New York: Carus Monographs No. 11, John Wiley & Sons, 1956.
- Paley, H. and P. M. Weichsel, *A First Course in Abstract Algebra*. New York: Holt, Rinehart and Winston, 1966.
- Passman, D. S., *Permutation Groups*. New York: W. A. Benjamin, 1969.
- Perlis, S., *Introduction to Algebra*. Waltham, Massachusetts: Blaisdell, 1966.
- Polites, G. W., *An Introduction to the Theory of Groups*. Scranton: International Text Book Company, 1968.
- Postnikov, M. M., *Foundations of Galois Theory*. Oxford: Pergamon Press, 1962.
- Redei, L., *Algebra*. Vol. I. Oxford: Pergamon Press, 1967.
- Ribenboim, P., *Rings and Modules*. New York: Interscience, 1969.
- Rose, John S., *A Course on Group Theory*. Cambridge: Cambridge University Press, 1978.
- Rotman, J. J., *Theory of Groups*. Boston: Allyn & Bacon, 1965.
- Sah, C. H., *Abstract Algebra*. New York: Academic Press, 1967.
- Samuel, P., *Algebraic Theory of Numbers*. London: Kershaw Publishing Company Ltd., 1972.
- Saracino Dam, *Abstract Algebra, a First Course*. Reading, Massachusetts: Addison-Wesley, 1980.
- Satake, J., *Linear Algebra*. New York: Marcel Dekker, 1975.
- Schwerdtfeger, H., *Introduction to Group Theory*. Leyden: No-ordhoff International Publishing, 1976.
- Scott, W. R., *Group Theory*. Englewood Cliffs, New Jersey: Prentice-Hall, 1964.
- Shapiro, L. W., *Introduction to Abstract Algebra*. New York: McGraw-Hill Book Company, 1975.
- Shephard, G. C., *Vector Spaces of Finite Dimension*. London: Oliver & Boyd, 1966.

- Sigler, L. E., *Algebra*. New York: Springer-Verlag, 1976.
- Smirnov, V. I., *Linear Algebra and Group Theory*. New York: Dover, 1961.
- Stewart, I., *Galois Theory*. London: Chapman and Hall, 1973.
- Van Der Waerden, B. L., *Modern Algebra*. Vol. I. New York: Ungar, 1949.
- , *Modern Algebra*. Vol. II. New York: Ungar, 1950.
- Wallace, D. A. R., *Groups*. London: George Allen & Unwin Ltd., 1974.
- Warner, S., *Classical Modern Algebra*. Englewood Cliffs, New Jersey: Prentice-Hall, 1971.
- , *Modern Algebra*. (Two Volumes). Prentice-Hall, 1965.
- Weilandt, H., *Finite Permutation Groups*. New York: Academic Press, 1964.
- Zariski, O. and P. Samuel, *Commutative Algebra*. Vol. I. Princeton, New Jersey: Von Nostrand, 1958.
- , *Commutative Algebra* Vol. II. Van Nostrand, 1960.
- Zassenhaus, H., *The Theory of Groups*. New York: Chelsea, 1958.

INDEX

- Abel 416
- Abelian group 42
- Absolutely algebraic 374
- Addition modulo n 44
- Admissible sequence 426
- Algebra 331
- Algebraically closed 361, 371
- Algebraic closure 361
- Algebraic element 357
- Algebraic extension 357
- Algebraic integer 374
- Algebraic number 369
- Alternating group, A_n 95
- Angle, trisection of 428
- Annihilator 340
- Annihilator ideal 458
- Artin 392
- Artinian ring 437
 - left 437
 - right 437
- Ascending chain condition (ACC) 433
- Associate 243
- Associative binary composition 25
- Automorphism
 - Inner 105
 - of a group 79
 - outer 177
- Basis 289
 - dual 339
 - ordered 304
 - orthonormal 316
 - standard 306
- Bessel's inequality 319
- Bezout's identity 251
- Bidual space 339
- Bijection 10
- Binary composition 24
- Boolean ring 272
- Butterfly Theorem 159
- Cancellation laws 50, 184
 - left 50, 184
 - right 50, 184
- Cartesian Product 5
- Cauchy-Schwarz inequality 312
- Cauchy's Theorems 128, 130
- Cayley-Hamilton Theorem 496
- Cayley's Theorem 95
- Centralizer 72
- Centre of a group 72
 - of a ring 198
- Characteristic
 - finite 352
 - zero 352
- Characteristic polynomial 468, 492
- Characteristic subgroup 108
- Characteristic value 468
- Characteristic vector 468
- Class
 - conjugate 111
 - equivalence 16
- Class equation 113
- Classical canonical matrix 473
- Classical quotient ring 231
- Co-domain 9
- Coefficient 234
- Column rank 347
- Column space 347
- Co-maximal ideals 204
- Commutative binary composition 24
- Commutative ring 180
- Commutator 152
- Commutator subgroup 152
- Companion matrix 463
- Complement of a subspace 299
- Composite number 34
- Composition series 158
- Congruence
 - modulo a subgroup, left 62
 - modulo a subgroup, right 61
 - modulo a subring, left 194
 - modulo a subring, right 194
 - modulo n 15

- Conjugacy 111
- Conjugate
 - class 111
 - elements over a field 366
 - space 338
- Constructible
 - circle 420
 - line 420
 - number 422
 - point 420
- Construction by ruler (straight edge)
 - and compass 420
- Co-prime elements 251
- Co-prime integers 34
- Correspondence 9
 - one-to-one (1-1) 9
- Correspondence Theorem
 - For groups 83
 - For rings 216
- Coset
 - left—, 62
 - right —, 62
- Countable set 17
- Cycle 90
- Cycles of a permutation 92
- Cyclic group 65, 106
- Cyclic permutation 90
- Cyclic space 465
- Cyclic subgroup 66
- Cyclotomic polynomial 404

- Degree
 - of algebraic element 358
 - of an extension 355
 - of a polynomial 238
- De Morgan's Laws 6
- Denumerable set 17
- Dependent linearly 287
- Derivative 375
- Derived group 152
- Descending chain condition (DCC) 437
- Diagonalizable 497
- Dihedral group 171
- Dimension 290
- Direct factor 124
- Direct product of groups
 - external 120
 - internal 122
- Direct summand 124
- Direct sum of groups
 - external 120
 - internal 122
- Direct sum of rings 443
- Direct sum of subspaces 299
- Distance 311
- Distributive law
 - left 179
 - right 179
- Divisible 240
- Divisible group 284
- Division algorithm
 - in integers 31
 - in $R[x]$ 239
- Division ring 183
- Domain 9
 - Euclidean 253
 - Gaussian 256
 - Integral 182
 - Principal Ideal 2.6
 - Unique Factorization 256
- Double Coset 135
- Dual
 - basis 339
 - space 338
- Eigen
 - value 449
 - vector 449
 - space 449
- Eisenstein's Criterion 266
- Element
 - algebraic 357
 - conjugate (in a group) 111
 - conjugate (over a field) 366
 - identity 25
 - inseparable 377
 - irreducible 244
 - order of 65
 - prime 244
 - primitive 378
 - separable 377
 - transcendental 218
- Elementary divisors 483
- Elementary symmetric function 415
- Embedding theorems for rings 218
- Empty set 2
- Endomorphism
 - of groups 77
 - of rings 213

- Epimorphism
 - of groups 77
 - of rings 213
- Equality of functions 9
- Equipotent sets 17
- Equivalence class 16
- Equivalence relation 15
- Equivalent linear transformations 346
- Equivalent matrices 346
- Equivalent, sub-normal series 158
- Euclidean
 - algorithm 32
 - domain 253
 - space 310
 - valuation 253
- Euler 74
- Euler's totient function
(Euler's ϕ -function) 36
- Even permutation 95
- Exponent 145
- Extension
 - algebraic 357
 - degree of 355
 - field 355
 - finite 356
 - Galois 394
 - inseparable 377
 - normal 397
 - radical 410
 - separable 377
 - simple 357
- Factor 243
- Factor group 76
- Factor groups of subnormal series 154
- Factorization Domain 272
- Factor theorem 362
- FC-group 117
- Fermat 74
- Field 183
- Field Extension 355
- Field of quotients 223
- Finite set 17
- Fixed field 391
- Fletcher's example 594
- Freshman's theorem 83
- Function 9
 - Invertible 13
- Fundamental theorem of
 - algebra 514
 - arithmetic 35
 - finite Abelian groups 145
 - finitely generated Abelian groups 501
 - Galois Theory 400
 - group homomorphism 81
 - ring homomorphism 215
 - vector space homomorphism 327
- Galois extension 394
- Galois field 380
- Galois group 390
- Galois theory
 - fundamental theorem of 400
- Gaussian domain 256
- Gaussian integer 190
- Gauss' theorem of
 - primitive polynomial 261
- Generalized ring of quotients 231
- Generator of a cyclic group 65
- Generic polynomial, n th 416
- Geometric multiplicity 450
- Gram-Schmidt Orthogonalization Process 316
- Group 42
 - decomposable 127
 - divisible 284
 - nilpotent 164
 - perfect 165
 - periodic 127
 - permutation 88
 - quaternion 58
 - quotient 76
 - simple 63
 - solvable 154
 - transformation 88
- Group ring 191
- Highest Common Factor (HCF) 32, 248
- Hilbert's Basis Theorem 436
- Homomorphism
 - group 77
 - ring 213
 - vector space 323
- Hua 227
- Ideal 195
 - finitely generated 201
 - left 195

- maximal 210
- minimal 212
- nil 202
- nilpotent 202
- primary 218
- prime 209
- principal 201
- right 195
- semi-prime 218
- Identity 25
 - left 25
 - mapping 10
 - right 25
- Image 9
- Inclusion map 213
- Independent, linearly 287
- Independent subspaces 303
- Indexed set 14
- Index of a subgroup 65
- Index of nilpotence 476
- Induced linear transformation 470
- Infinite set 17
- Inner automorphism 105
- Inner product space 309
- Integer, partition 114
- Integral domain 182
- Intersection of sets 2
- Invariants of a nilpotent transformation 489
- Invariant subgroup 63
- Invariant subspace 456
- Inverse of a function 13
- Inverse of an element 25
 - left 25
 - right 25
- Invertible element 25
- Irreducible element 244
- Irreducible polynomial 262
- Isomorphism of groups 79
 - of rings 213
 - of vector spaces 323
- Jordan Block 469
- Jordan canonical form 473
- Jordan Holder theorem 162
- Jordan matrix 465
- Jordan normal form 486
- Kaplansky* 188
- Kernel of a group homomorphism 78
- Kernel of a ring homomorphism 214
- Klein's 4-group 48
- Kronecker* 363
- Lagrange* 64
- Leading coefficient of a polynomial 234
- Least common multiple (*LCM*) 33, 248
- Left coset 62
- Left inverse 25
- Left invertible 445
- Length of a cycle 90
- Linear combination 280
- Linear functional 338
- Linearly dependent (*L.D.*) 287
- Linearly independent (*L.I.*) 287
- Linear transformation 323
 - Cyclic 465
 - left invertible 445
 - nilpotent 488
 - non-singular 452
 - regular 445
 - right invertible 445
 - singular 452
- Mapping 9
 - composite 10
 - identity 10
 - inclusion 213
 - 1-1 9
 - onto 9
- Matrix relative to a basis 305
- Matrix units 270
- Maximal ideal 210
- Maximal normal subgroup 157
- Maximal subgroup 174
- Minimal ideal 212
- Minimal polynomial
 - of a linear transformation 447
 - of an element 358
 - of a matrix 447
- Modular ideal 273
- Modular law 205
- Monomorphism
 - of groups 77
 - of rings 213
- Moore* 381
- Motzkin's Example 506
- Multiplication modulo n 44

- Multiplicative subset 230
- Natural homomorphism
 - for groups 82
 - for rings 213
- Niven 518
- Noetherian ring 433
 - left 433
 - right 433
- Non-Singular linear transformation 452
- Norm 311
- Normal extension 397
- Normalizer 72
- Normal series 154
- Normal subgroup 63
 - Maximal 157
- Nullity of a linear transformation 325
- Null set 2
- Null space 325
- Number
 - algebraic 369
 - constructible 422
 - transcendental 518
- Odd permutation 95
- One-one or 1-1 mapping 9
- One to one correspondence 9
- Onto mapping 9
- Orbit 89
- Ordered pair 4
- Order of a group 64
- Order of an element 65
- Ore 231
- Orthogonal
 - complement 318
 - subset 314
 - vectors 313
- Orthonormal
 - basis 316
 - set 314
- Outer automorphism 177
- Over ring 193
- Partial ordering 15
- Partition 114
- Perfect field 382
- Perfect group 165
- Permutation 11
 - cyclic 90
 - disjoint 90
 - even 95
 - group 88
 - odd 95
 - similar 114
- p -group 130
- Polynomial 234
 - content of 260
 - cyclotomic 404
 - degree of 238
 - inseparable 377
 - irreducible 262
 - monic 357
 - primitive 260
 - root of 357
 - separable 376
- Poincare 69
- Poretsky's law 30
- Powers of an ideal 202
- Primary ideal 218
- Prime element 244
- Prime field 351
- Prime ideal 209
- Prime number 34
- Prime subfield 352
- Primitive element 378
- Primitive n th root of unity 403
- Primitive polynomial 260
- Principal ideal 201
- Principal ideal domain (PID) 246
- Principal ideal ring 246
- Principle of duality 339
- Product ideal 510
- Projection 335
 - supplementary orthogonal 335
- Proper subset 2
- Pseudo ring 188
- p -subgroup 130
- Quaternion group 56, 172
- Quotient
 - group 76
 - ring 208
 - space 297
- Radical Extension 410
- Range of a function 9
- Rank
 - column 347

- of a linear transformation 325
- row 397
- Rational canonical form 491
- Refinement 154
 - proper 154
- Reflexive relation 15
- Regular
 - element 230
 - linear transformation 446
 - ring 273
 - semi-group 56
- Relation, binary 9
- Relatively prime elements 257
- Relatively prime integers 34
- Remainder 240
- Restriction 18
- Right coset 82
- Right inverse 25
- Right invertible 445
- Ring 179
 - commutative 180
 - division 183
 - of quotients 230
 - regular 273
 - valuation 272
 - with unity 180
- Root
 - multiple 375
 - multiplicity of 362
 - of a polynomial 357
- Row rank 347
- Row space 347
- Schreier's Refinement theorem 160
- Schroder-Bernstein's theorem 19
- Schwarz inequality 312
- Scot's example 504
- Self-conjugate subgroup 63
- Semi-group 43
 - ring 186
- Semi-prime ideal 218
- Separable
 - element 377
 - extension 377
 - polynomial 376
- Set
 - empty 2
 - orthogonal 314
- Side divisor 512
- Similar
 - linear transformation 446
 - matrices 346
 - permutations 114
- Simple
 - extension 357
 - group 61
 - ring 183
- Singular linear transformation 452
- Solvable by radicals 410
- Solvable group 154
- Solvable series 154
- Space
 - bidual 339
 - conjugate 338
 - dual 338
 - eigen 449
 - Euclidean 310
 - quotient 297
 - unitary 310
 - vector 276
- Splitting field 364
- Standard inner product 309
- Steinitz theorem 384
- subfield 350
- Subgroup 59
 - generated by a subset 151
 - maximal 174
 - normal 63
- Subnormal series 154
- Subring 193
- Subset 2
 - multiplicative 230
- Subspace 278
- Sum
 - of ideals 201
 - of subspace 281
- Supplementary orthogonal projections 335
- Surjection 9
- Sylow's p -group 131
- Sylow's theorems 129, 132
- Sylvester law of Nullity 326
- Symmetric
 - difference 30
 - function 415
 - group 95

- relation 15
- T -invariant 470
- Total quotient ring 231
- Transcendence of e 518
- Transcendental element 233
- Transcendental number 518
- Transformation group 88
- Transitive relation 15
- Transpose of a linear transformation 343
- Transposition 91
- Union of sets 2
- Unique factorization domain (UFD) 256
- Unitary space 310
- Unit in a ring 183
- Universal side divisor 512
- Valuation ring 272
- Vector space 276
 - complex 309
 - real 309
- Wedderburn's Theorem 185
- Wilson's Theorem 176
- Zassenhaus* 159
- Zero 179
- Zero-divisor
 - left 182
 - right 182

MODERN ALGEBRA

Surjeet Singh

Qazi Zameeruddin

The book has been thoroughly revised and alternative proofs of Sylow's Theorems using double-coset technique have been added. Many solved exercises have been added to give ample guidance in understanding the underlying abstract notions. Problems are provided with sufficient hints and an appendix is given at the end of the book which includes fundamental theorem of finitely generated Abelian Groups, Fundamental Theorem of Algebra and Transcendence of 'e', and some examples of difficult problems which are not easily accessible.

The book starts from Set Theory and covers a sufficiently advanced course in Group Theory and Ring Theory. A detailed study of Field Theory and its application to Geometry is undertaken after a brief and concise account of Vector Spaces and Linear Transformations. The last chapter discusses rings with chain conditions and Hilbert's famous theorem.

SURJEET SINGH is Professor of Mathematics at University of Kuwait, Kuwait. After getting his Ph.D. from University of Delhi, he worked as a lecturer in K.M. College, University of Delhi as reader in Aligarh Muslim University and then as Professor at Guru Nanak Dev University, Amritsar. He has a number of research papers to his credit and is presently doing research in Torsion Theories.

QAZI ZAMEERUDDIN is lecturer of Mathematics at K.M. College, University of Delhi. He got his Ph.D. in Ring Theory from University of Delhi. His teaching span is of 21 years. He is co-author of *Business Mathematics* and *Solid Geometry* published by VIKAS. He is mainly interested in theory of groups, rings, modules and fields.

1V2SJ610

ISBN 0-7069-4216-7

Rs 55